

# Smooth plug-in inverse estimators in the current status continuous mark model.

P. Groeneboom<sup>a</sup>, G. Jongbloed<sup>a</sup> & B.I. Witte<sup>b</sup>

<sup>a</sup>*Delft University of Technology*; <sup>b</sup>*VU University Medical Center*

January 12, 2013

**ABSTRACT.** We consider the problem of estimating the joint distribution function of the event time and a continuous mark variable when the event time is subject to interval censoring case 1 and the continuous mark variable is only observed in case the event occurred before the time of inspection. The nonparametric maximum likelihood estimator in this model is known to be inconsistent. We study two alternative smooth estimators, based on the explicit (inverse) expression of the distribution function of interest in terms of the density of the observable vector. We derive the pointwise asymptotic distribution of both estimators.

*Key words:* asymptotic distribution, bivariate kernel estimation, continuous mark variable, consistency, current status data, plug-in estimation

## 1 Introduction

To test the efficacy of a vaccine, preventative trials are held where participants are injected with the vaccine and tested for several times. One of the questions of interest in the trials is whether the efficacy depends on the genetic sequence of the exposing virus. To answer this question, Flynn et al. (2005) studied the so-called viral distance between the HIV sequence represented in the vaccine and the HIV sequence the participant is infected with. This distance can be considered as a “mark” variable, since it can only be observed if infection has already taken place. This variable is possibly correlated with the time of HIV infection and according to Gilbert et al. (2001) it is natural to treat it as a continuous random variable.

A natural statistical model to describe the observations in these HIV vaccine trials is the interval censored continuous mark model, which was first studied by Hudgens et al. (2007). In this model,  $X$  is an event time (the time of HIV infection) and  $Y$  is a continuous mark variable

(the viral distance) and we are interested in the bivariate distribution function  $F_0$  of the pair  $(X, Y)$ . However, the event time is subject to interval censoring case  $k$ . We restrict ourselves to the special instance of interval censoring case 1 (also known as current status censoring) and refer to this model as the current status continuous mark model.

For this model, the method of nonparametric maximum likelihood estimation is studied Maathuis and Wellner (2008). There it is proved that the maximum likelihood estimator (MLE) is inconsistent. An approach they propose to ‘repair’ the inconsistency is by discretizing the mark variable. Discretizing the mark variable to  $K$  levels, the resulting observations can be viewed as observations from the current status  $K$ -competing risk model. The characterization, consistency and (local) asymptotic distribution theory of the MLE in that model follow from Groeneboom et al. (2008a, 2008b). Results on consistency and asymptotics as  $K \rightarrow \infty$  are not yet known.

Another natural way to estimate the distribution function  $F_0$  is by viewing this problem as an inverse statistical model. In inverse models, like interval censoring models or deconvolution models, one is interested in estimating the distribution of a random variable  $X$ . Instead of observing this variable  $X$  directly, only a related variable  $W$  is observed. The distribution of  $W$  depends on the distribution function  $F_0$  of  $X$  (or its Lebesgue density  $f_0$ ) via a known (direct) relation. In some cases, this relation can be explicitly inverted to express  $F_0$  in terms of the distribution of  $W$ , and to estimate  $F_0$  one can plug in an estimator for the distribution of  $W$  in this inverse relation. The resulting estimator is called a plug-in inverse estimator. Plug-in inverse estimators are studied by Hall and Smith (1988) in Wicksell’s corpuscle problem, by Stefanski and Carroll (1990) in the deconvolution model and by Burke (1988) in the bivariate right-censoring model.

In this paper we study plug-in inverse estimators in the current status continuous mark model. We start with a formal description of the model and define two plug-in inverse estimators in Section 2. One estimator is based on univariate kernel smoothing, the other is based on bivariate kernel smoothing. In Section 3, we prove that these estimators are uniformly consistent for  $F_0$ . Unfortunately, these estimators are not monotonically increasing in both directions, which is a necessary property of bivariate distribution functions. In Section 3 we prove that the estimator based on bivariate kernel smoothing asymptotically will have all properties of a bivariate distribution function on a large subset of  $[0, \infty)^2$ . The plug-in inverse estimator resulting from the univariate kernel smoothing estimator is computationally and asymptotically more tractable. In Section 4, we first derive the asymptotic distribution of this estimator. After that, we prove that for certain choices of the smoothing parameter in the  $z$ -direction, the two plug-in inverse estimators are asymptotically equivalent, while for other choices the asymptotic

biases differ but the asymptotic variances are equal. This phenomenon was also observed by Marron and Padgett (1987) and Patil et al. (1994) in the case of estimating densities based on right-censored data and by Groeneboom et al. (2010) in the current status model. The asymptotic distribution of the estimator based on bivariate kernel smoothing then follows easily. In Section 5, we briefly address the problem of estimating smooth functionals. A small simulation study to compare the estimators with the binned MLE studied by Maathuis and Wellner (2008) and the maximum smoothed likelihood estimator studied by Groeneboom et al. (2010) is performed in Section 6. Technical proofs and lemmas can be found in the Appendix.

## 2 Definition of the estimators

In this section we describe the current status continuous mark model in more detail and define two plug-in inverse estimators based on kernel smoothing.

Let  $X$  be an event time,  $Y$  a continuous mark variable and  $F_0$  be the distribution function of the pair  $(X, Y)$ . In the current status continuous mark model, instead of observing the pair  $(X, Y)$ , we observe a censoring variable  $T$ , independent of  $(X, Y)$  with Lebesgue density  $g$ , as well as the indicator variable  $\Delta = 1_{\{X \leq T\}}$ . In case  $X \leq T$ , i.e. if  $\Delta = 1$ , we also observe the mark variable  $Y$ ; in case  $\Delta = 0$  the variable  $Y$  is not observed. Under the assumption that  $P(Y = 0) = 0$ , we can represent the observable information on  $(X, Y)$  in the vector  $W = (T, Z, \Delta)$ , for  $Z = \Delta \cdot Y$ .

Let  $\lambda_i$  be Lebesgue-measure on  $\mathbb{R}^i$ ,  $\mathcal{B}$  the Borel  $\sigma$ -algebra on  $[0, \infty)^2$  and define the measure  $\lambda$  on  $\mathcal{B}$  by

$$\lambda(B) = \lambda_2(B) + \lambda_1(\{x \in [0, \infty) : (x, 0) \in B\}), \quad B \in \mathcal{B}.$$

Then, the density of the observable vector  $W$  w.r.t. the product of this measure with counting measure on  $\{0, 1\}$  can be written as

$$h_{F_0}(t, z, \delta) = \delta g(t) \partial_2 F_0(t, z) + (1 - \delta) g(t) (1 - F_{0,X}(t)) = \delta h_1(t, z) + (1 - \delta) h_0(t), \quad (1)$$

where  $F_{0,X}$  is the marginal distribution of  $X$  and  $\partial_2 F_0(t, z) = \frac{\partial}{\partial z} F_0(t, z)$ . More generally, for convenience of notation, we denote the  $j$ th partial derivative with respect to  $x_i$  of a function  $F$  by  $\partial_i^j F$ , i.e.

$$\partial_i^j F(x_1, x_2) = \frac{\partial^j}{\partial y_i^j} F(y_1, y_2) \Big|_{(y_1, y_2) = (x_1, x_2)},$$

and omit  $j$  when  $j = 1$ .

Based on the relation  $h_1(t, z) = g(t)\partial_2 F_0(t, z)$ , we can express the bivariate distribution function  $F_0$  of  $(X, Y)$  in terms of the (sub-)densities  $g$  and  $h_1$

$$F_0(t, z) = \frac{1}{g(t)} \int_0^z h_1(t, v) dv. \quad (2)$$

Then, our plug-in inverse estimator in the current status continuous mark model is defined as

$$\hat{F}(t, z) = \frac{1}{\hat{g}(t)} \int_0^z \hat{h}_1(t, v) dv,$$

where  $\hat{g}$  and  $\hat{h}_1$  are estimators for  $g$  and  $h_1$ , respectively.

Before explicitly choosing the estimators  $\hat{g}$  and  $\hat{h}_1$ , we introduce some notation. Throughout the paper  $k$  denotes a univariate kernel density,  $\tilde{k}$  a bivariate kernel density and  $(\alpha_n)$  and  $(\beta_n)$  vanishing sequences of positive smoothing parameters. Let  $k_{\alpha_n}$  and  $\tilde{k}_{\alpha_n, \beta_n}$  the rescaled versions of  $k$  and  $\tilde{k}$ , i.e.,  $k_{\alpha_n}(u) = \alpha_n^{-1}k(u/\alpha_n)$  and  $\tilde{k}_{\alpha_n, \beta_n}(u, v) = \alpha_n^{-1}\beta_n^{-1}\tilde{k}(u/\alpha_n, v/\beta_n)$ . Furthermore, we define

$$m_2(k) = \int u^2 k(u) du, \quad m_2(\tilde{k}) = \iint w_1^2 \tilde{k}(w_1, w_2) dw_1 dw_2.$$

Then for fixed  $t_0$  and  $z_0$ , we estimate  $g$  and  $h_1$  by their respective univariate and bivariate kernel (sub-)density estimators

$$\hat{g}_n(t_0) = \frac{1}{n} \sum_{i=1}^n k_{\alpha_n}(t_0 - T_i), \quad \hat{h}_{n,1}^{(2)}(t_0, z_0) = \frac{1}{n} \sum_{i=1}^n \Delta_i \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z_0 - Z_i).$$

The plug-in inverse estimator then becomes

$$\hat{F}_n^{(2)}(t_0, z_0) = \frac{\int_0^{z_0} \frac{1}{n} \sum_{i=1}^n \Delta_i \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz}{\frac{1}{n} \sum_{i=1}^n k_{\alpha_n}(t_0 - T_i)}. \quad (3)$$

Here, superscript 2 in the notation for the plug-in estimator refers to the fact that there is smoothing in two directions.

In Section 4 we also consider a less natural, but computationally and asymptotically more tractable estimator using an estimate for the numerator  $\int_0^{z_0} h_1(t_0, z) dz$  based on smoothing only in the  $t$ -direction, i.e., when we estimate it by

$$\frac{1}{n} \sum_{i=1}^n 1_{[0, z_0]}(Z_i) \Delta_i k_{\alpha_n}(t_0 - T_i).$$

The plug-in inverse estimator then becomes

$$\hat{F}_n^{(1)}(t_0, z_0) = \frac{\frac{1}{n} \sum_{i=1}^n 1_{[0, z_0]}(Z_i) \Delta_i k_{\alpha_n}(t_0 - T_i)}{\frac{1}{n} \sum_{i=1}^n k_{\alpha_n}(t_0 - T_i)}. \quad (4)$$

Superscript 1 in the notation for this estimator refers to the fact that there is only smoothing in one direction. Note that if we take  $k(y) = \frac{1}{2}1_{[-1, 1]}(y)$ , (4) results in

$$\hat{F}_n^{(1)}(t_0, z_0) = \frac{\int_{u \in A_n} \int_{z \leq z_0} \delta d\mathbb{H}_n(u, z, \delta)}{\int_{u \in A_n} \int_{z \geq 0} d\mathbb{H}_n(u, z, \delta)},$$

where  $\mathbb{H}_n$  is the empirical distribution of the observations  $(T_1, Z_1, \Delta_1), \dots, (T_n, Z_n, \Delta_n)$  and  $A_n = A_n(t_0) = [t_0 - \alpha_n, t_0 + \alpha_n]$ . This estimator is the total number of observations  $T_i$  in  $A_n$  with  $Z$ -value smaller than or equal to  $z_0$  and  $\Delta = 1$  divided by the total number of observations  $(T_i, Z_i)$  in the strip  $A_n \times [0, \infty)$ .

It is very natural to define the kernel density  $k$  in terms of the kernel density  $\tilde{k}$  as stated in assumption (K.1):

(K.1) Let  $\tilde{k}$  be a bivariate kernel density, then the kernel density  $k$  is defined as

$$k(w_1) = \int \tilde{k}(w_1, w_2) dw_2.$$

Indeed, if (K.1) holds the estimator  $\hat{F}_n^{(2)}$  also satisfies the inverse relation  $h_0(t) = g(t)(1 - F_{0,X}(t))$  that follows from substituting  $\delta = 0$  in (1). To see this, note that we have that

$$\begin{aligned} \hat{g}_n(t_0) &= \frac{1}{n} \sum_{i=1}^n k_{\alpha_n}(t_0 - T_i) = \frac{1}{n} \sum_{i=1}^n (1 - \Delta_i) k_{\alpha_n}(t_0 - T_i) + \frac{1}{n} \sum_{i=1}^n \Delta_i k_{\alpha_n}(t_0 - T_i) \\ &= \frac{1}{n} \sum_{i=1}^n (1 - \Delta_i) k_{\alpha_n}(t_0 - T_i) + \int \frac{1}{n} \sum_{i=1}^n \Delta_i \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz. \end{aligned}$$

If we define  $\hat{h}_{n,0}(t_0) = \frac{1}{n} \sum_{i=1}^n (1 - \Delta_i) k_{\alpha_n}(t_0 - T_i)$  as an estimator for the sub-density  $h_0$  in (1), then

$$1 - \hat{F}_{n,X}^{(2)}(t_0) = 1 - \hat{F}_n^{(2)}(t_0, \infty) = 1 - \frac{\int_0^\infty \hat{h}_{n,1}(t_0, z) dz}{\hat{g}_n(t_0)} = 1 - \frac{\hat{g}_n(t_0) - \hat{h}_{n,0}(t_0)}{\hat{g}_n(t_0)} = \frac{\hat{h}_{n,0}(t_0)}{\hat{g}_n(t_0)}.$$

Figure 1 illustrates the estimator  $\hat{F}_n^{(1)}$  for  $n = 10$  and  $n = 100$ . For  $F_0$  we took the uniform distribution on  $[0, 1]^2$  and for  $g$  the uniform distribution on  $[0, 1]$ . As kernel density we used  $k(y) = \frac{1}{2} 1_{[-1, 1]}(y)$ . The smoothing parameter  $\alpha_n$  is taken to be 0.65 for  $n = 10$  and 0.40 for  $n = 100$ . These values are chosen for illustrative purpose only and do not depend on the data. In Section 7 we briefly address the problem of choosing  $\alpha_n$  and  $\beta_n$  depending on the data.

[Figure 1 here]

Note that these estimators are not true bivariate distribution functions, as they decrease locally in the  $x$ -direction. Monotonicity of a bivariate function is a necessary (but not sufficient) condition in order to be a bivariate distribution function, hence these estimators can be seen as naive estimators. The estimator  $\hat{F}_n^{(2)}$  can also have this undesirable naive behavior.

### 3 Consistency and monotonicity

In this section we prove that the estimators  $\hat{F}_n^{(1)}$  and  $\hat{F}_n^{(2)}$  are uniformly consistent. Furthermore, we prove that for appropriate choices of the bandwidths and  $n$  sufficiently large,  $\hat{F}_n^{(2)}$  will

have all properties of a bivariate distribution function on a large subset of  $[0, \infty)^2$ , with arbitrarily high probability. To derive these results for  $\hat{F}_n^{(2)}$ , we assume the distribution function of interest  $F_0$  and the censoring density  $g$  satisfy the following conditions.

(F.1) The Lebesgue density  $f_0$  of  $F_0$  exists for all  $(t, z) \in [0, \infty)^2$ .

(G.1) Let  $\mathcal{S}_{0,X}^\circ$  denote the interior of the support of the marginal density  $f_{0,X}$  of  $X$ . On  $\mathcal{S}_{0,X}^\circ$ , the density  $g$  satisfies  $0 < g < \infty$  and its derivative  $g'$  is uniformly continuous and bounded.

We also impose some conditions on the kernel densities  $k$  and  $\tilde{k}$ , as well as a condition on the smoothing parameters  $\alpha_n$  and  $\beta_n$ .

(K.2) The kernel density  $k$  has compact support  $[-1, 1]$ , is continuous and symmetric around 0.

(K.3) The kernel density  $\tilde{k}$  has compact support  $[-1, 1]^2$ , is continuous and satisfies

$$\iint w_i \tilde{k}(w_1, w_2) dw_1 dw_2 = 0 \quad (i = 1, 2), \quad \iint w_2^2 \tilde{k}(w_1, w_2) dw_1 dw_2 = \iint w_1^2 \tilde{k}(w_1, w_2) dw_1 dw_2.$$

(C.1) The positive smoothing parameters  $\alpha_n$  and  $\beta_n$  satisfy

$$\lim_{n \rightarrow \infty} \alpha_n = \lim_{n \rightarrow \infty} \beta_n = 0, \quad \lim_{n \rightarrow \infty} n\alpha_n = \infty.$$

A possible choice for the bivariate kernel density  $\tilde{k}$  is the product kernel density  $\tilde{k}(x, y) = k_1(x)k_2(y)$  for univariate kernel densities  $k_1$  and  $k_2$  with compact support  $[-1, 1]$  that are continuous and symmetric around 0. This kernel density  $\tilde{k}$  satisfies condition (K.1) for  $k = k_1$  and (K.3) if  $m_2(k_1) = m_2(k_2)$ .

**Theorem 1** *Assume  $F_0$  and  $g$  satisfy conditions (F.1) and (G.1). Also assume  $k$  is defined via relation (K.1) and satisfies condition (K.2). Furthermore, let  $\alpha_n$  and  $\beta_n$  satisfy condition (C.1). Let  $\mathcal{A} \subset \mathbb{R}_+^2$  be a compact set such that  $g(t) \geq c > 0$  for all  $(t, z) \in \mathcal{A}$ . Then  $\hat{F}_n^{(1)}$  and  $\hat{F}_n^{(2)}$  are uniformly consistent on  $\mathcal{A}$ .*

*Proof:* The uniform consistency of  $\hat{F}_n^{(1)}$  follows from Theorem 3.2 in Härdle et al. (1988).

To prove that  $\hat{F}_n^{(2)}$  is uniformly consistent on  $\mathcal{A}$ , first note that for  $n$  sufficiently large there exists  $\varepsilon > 0$  such that

$$\sup_{(t,z) \in \mathcal{A}} |\hat{h}_{n,1}^{(2)}(t, z) - h_1(t, z)| \leq \varepsilon,$$

see also Lemma 8. Hence

$$\left| \int_0^z \hat{h}_{n,1}^{(2)}(t, y) dy - \int_0^z h_1(t, y) dy \right| \leq \int_0^z |\hat{h}_{n,1}^{(2)}(t, y) - h_1(t, y)| dy \leq \varepsilon z.$$

Since  $z \in \mathcal{A}$  and  $\mathcal{A}$  is compact, this implies that

$$\sup_{(t,z) \in \mathcal{A}} \left| \int_0^z \hat{h}_{n,1}^{(2)}(t, y) dy - \int_0^z h_1(t, y) dy \right| \xrightarrow{\mathcal{P}} 0. \quad (5)$$

Write  $N_n^{(2)}(t, z) = \int_0^z \hat{h}_{n,1}^{(2)}(t, y) dy$  and  $N(t, z) = \int_0^z h_1(t, y) dy$ . Then we have that

$$\begin{aligned} |\hat{F}_n^{(2)}(t, z) - F_0(t, z)| &= \left| \frac{N_n^{(2)}(t, z)}{\hat{g}_n(t)} - \frac{N(t, z)}{g(t)} \right| \\ &\leq \left| \frac{N_n^{(2)}(t, z) - N(t, z)}{g(t)} \right| + \left| \frac{N_n^{(2)}(t, z)}{\hat{g}_n(t)} - \frac{N_n^{(2)}(t, z)}{g(t)} \right| \\ &= \frac{1}{g(t)} |N_n^{(2)}(t, z) - N(t, z)| + N_n^{(2)}(t, z) \left| \frac{1}{\hat{g}_n(t)} - \frac{1}{g(t)} \right|. \end{aligned}$$

The first term converges uniformly to zero in probability over  $\mathcal{A}$  by (5). The second term converges uniformly to zero in probability by Lemma 8, and uniform consistency of  $\hat{F}_n^{(2)}$  follows.  $\square$

Each bivariate distribution function  $F$  has to satisfy

$$\forall x_1 < x_2, y_1 < y_2 : F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geq 0. \quad (6)$$

This condition requires that each rectangle  $[x_1, x_2] \times [y_1, y_2]$  has a nonnegative mass and suggests that some shape constraints on  $F_0$  are imposed by the model. However, in Theorem 2 below, we prove that it is not necessary to use this shape constraint to estimate  $F_0$  since the estimator  $\hat{F}_n^{(2)}$  satisfies condition (6) asymptotically. To prove this, we prove that the Lebesgue density  $\hat{f}_n^{(2)}$  is positive, with probability converging to one. The estimator  $\hat{F}_n^{(1)}$  does not have a density w.r.t. Lebesgue measure  $\lambda_2$ , hence a similar result can not be proved in this way for  $\hat{F}_n^{(1)}$ . To prove Theorem 2, we need stronger conditions on  $\alpha_n$  and  $\beta_n$  than condition (C.1).

(C.2) The smoothing parameters  $\alpha_n$  and  $\beta_n$  converge to zero as  $n \rightarrow \infty$  and satisfy

$$\lim_{n \rightarrow \infty} n\alpha_n^2\beta_n^2 = \infty, \quad \lim_{n \rightarrow \infty} n\alpha_n^3\beta_n = \infty.$$

Note that sequences  $\alpha_n$  and  $\beta_n$  satisfying condition (C.2) also satisfy condition (C.1) and  $n\alpha_n^3 \rightarrow \infty$ .

**Theorem 2** Assume  $F_0$  and  $g$  satisfy conditions (F.1) and (G.1). Also assume  $k$  and  $\tilde{k}$  satisfy conditions (K.2) and (K.3). In addition, assume  $k'$  and  $\partial_1 \tilde{k}$  are uniformly continuous. Furthermore, let  $\alpha_n$  and  $\beta_n$  satisfy condition (C.2). Let  $\mathcal{S} \subset [0, \infty)^2$  be compact and such that  $f_0$  is uniformly continuous on an open subset that contains  $\mathcal{S}$  and for all  $\delta > 0$ ,  $\mathcal{M}_\delta = \{(t, z) \in [0, \infty)^2 : f_0(t, z) \geq 2\delta\} \cap \mathcal{S}$ . Then for  $\delta > 0$ ,

$$P \left( \forall (t, z) \in \mathcal{M}_\delta : \hat{f}_n^{(2)}(t, z) > \frac{l_g^2}{2u_g^2} \delta \right) \rightarrow 1, \quad (7)$$

where  $\hat{f}_n^{(2)}$  is the Lebesgue density of  $\hat{F}_n^{(2)}$  and  $l_g$  and  $u_g$  are as defined in Lemma 7.

*Proof:* Fix  $\delta > 0$ . First note that since

$$\frac{\partial^2}{\partial t \partial z} \int_0^z \hat{h}_{n,1}^{(2)}(t, v) dv = \partial_1 \hat{h}_{n,1}^{(2)}(t, z)$$

we have the following expression for  $\hat{f}_n^{(2)}$

$$\hat{f}_n^{(2)}(t, z) = \frac{\partial^2}{\partial u \partial v} \hat{F}_n^{(2)}(u, v) \Big|_{(u,v)=(t,z)} = \frac{\hat{g}_n(t) \partial_1 \hat{h}_{n,1}^{(2)}(t, z) - \hat{g}'_n(t) \hat{h}_{n,1}^{(2)}(t, z)}{\hat{g}_n(t)^2}. \quad (8)$$

We first consider the numerator and prove that

$$P(\forall (t, z) \in \mathcal{M}_\delta : \hat{g}_n(t) \partial_1 \hat{h}_{n,1}^{(2)}(t, z) - \hat{g}'_n(t) \hat{h}_{n,1}^{(2)}(t, z) > 2l_g^2 \delta) \longrightarrow 1. \quad (9)$$

For this, note that for all  $(t, z) \in \mathcal{M}_\delta$

$$\begin{aligned} \hat{g}_n(t) \partial_1 \hat{h}_{n,1}^{(2)}(t, z) - \hat{g}'_n(t) \hat{h}_{n,1}^{(2)}(t, z) &= \hat{g}_n(t) (\partial_1 \hat{h}_{n,1}^{(2)}(t, z) - \partial_1 h_1(t, z)) + \hat{h}_{n,1}^{(2)}(t, z) (g'(t) - \hat{g}'_n(t)) \\ &\quad + \partial_1 h_1(t, z) (\hat{g}_n(t) - g(t)) + g'(t) (h_1(t, z) - \hat{h}_{n,1}^{(2)}(t, z)) + g(t) \partial_1 h_1(t, z) - g'(t) h_1(t, z) \\ &\geq - \sup_{t \in \text{proj}_X \mathcal{M}_\delta} \hat{g}_n(t) \sup_{(t,z) \in \mathcal{M}_\delta} |\partial_1 \hat{h}_{n,1}^{(2)}(t, z) - \partial_1 h_1(t, z)| \\ &\quad - \sup_{(t,z) \in \mathcal{M}_\delta} \hat{h}_{n,1}^{(2)}(t, z) \sup_{t \in \text{proj}_X \mathcal{M}_\delta} |g'(t) - \hat{g}'_n(t)| \\ &\quad - \sup_{(t,z) \in \mathcal{M}_\delta} \partial_1 h_1(t, z) \sup_{t \in \text{proj}_X \mathcal{M}_\delta} |\hat{g}_n(t) - g(t)| \\ &\quad - \sup_{t \in \text{proj}_X \mathcal{M}_\delta} g'(t) \sup_{(t,z) \in \mathcal{M}_\delta} |h_1(t, z) - \hat{h}_{n,1}^{(2)}(t, z)| + g(t) \partial_1 h_1(t, z) - g'(t) h_1(t, z), \end{aligned}$$

with  $\text{proj}_X \mathcal{M}_\delta = \{t : (t, z) \in \mathcal{M}_\delta \text{ for some } z\}$ . By Lemma 8 all random terms converge to zero in probability. Since  $g(t) \partial_1 h_1(t, z) - g'(t) h_1(t, z) = g(t)^2 f_0(t, z)$  we have that the last term is bounded below by  $\inf_{(t,z) \in \mathcal{M}_\delta} g(t)^2 f_0(t, z) \geq 2l_g^2 \delta$  by Lemma 7.

By Lemma 7 and the uniform consistency of  $\hat{g}_n$  [see Lemma 8], we have that  $0 < \frac{1}{2} l_g < \hat{g}_n(t) < 2u_g < \infty$  for all  $t \in \text{proj}_X \mathcal{M}_\delta$  with probability converging to one. This implies that for all  $(t, z) \in \mathcal{M}_\delta$

$$\hat{f}_n^{(2)}(t, z) \geq \frac{\hat{g}_n(t) \partial_1 \hat{h}_{n,1}^{(2)}(t, z) - \hat{g}'_n(t) \hat{h}_{n,1}^{(2)}(t, z)}{4u_g^2} > \frac{2l_g^2}{4u_g^2} \delta,$$

with probability converging to one. Hence (7) follows.  $\square$

*Remark.* If, in addition to condition (F.1), we assume that  $f_0$  is uniformly continuous on  $[0, \infty)^2$ , this theorem implies that for each  $\delta > 0$  and  $M > 0$ , the restriction of  $\hat{F}_n^{(2)}$  to the set  $\{(t, z) \in [0, M]^2 : f_0(t, z) \geq \delta\}$  will asymptotically be the restriction to this set of a bivariate distribution function  $\tilde{F}_n$  on  $[0, \infty)^2$ .



## 4 Asymptotic distributions

In this section we derive the asymptotic distribution of both plug-in inverse estimators. Although the estimator  $\hat{F}_n^{(2)}$  is more natural, we start with the estimator  $\hat{F}_n^{(1)}$  since deriving its asymptotic distribution is easier. Subsequently, we prove that for certain choices of the smoothing parameter  $\beta_n$  the estimators  $\hat{F}_n^{(1)}$  and  $\hat{F}_n^{(2)}$  are asymptotically equivalent, yielding the asymptotic distribution of  $\hat{F}_n^{(2)}$ .

**Theorem 3** *Assume  $F_0$  and  $g$  satisfy conditions (F.1) and (G.1). Also assume  $k$  satisfies condition (K.2). Fix  $t_0, z_0 > 0$  such that  $\partial_1^2 F_0(t, z)$  and  $g''(t)$  exist and are continuous in a neighborhood of  $(t_0, z_0)$  and  $t_0$ , respectively, and  $\partial_1^2 F_0(t_0, z_0) + 2g'(t_0)\partial_1 F_0(t_0, z_0)/g(t_0) \neq 0$  and  $g(t_0) > 0$ . Then for  $\alpha_n = cn^{-1/5}$ ,*

$$n^{2/5}(\hat{F}_n^{(1)}(t_0, z_0) - F_0(t_0, z_0)) \rightsquigarrow \mathcal{N}(\mu_1, \sigma^2)$$

where

$$\mu_1 = \frac{1}{2}c^2 m_2(k) \left\{ \partial_1^2 F_0(t_0, z_0) + 2 \frac{g'(t_0)\partial_1 F_0(t_0, z_0)}{g(t_0)} \right\}, \quad (10)$$

$$\sigma^2 = c^{-1} \frac{F_0(t_0, z_0)(1 - F_0(t_0, z_0))}{g(t_0)} \int k(u)^2 du. \quad (11)$$

*Remark.* In case  $\partial_1^2 F_0(t_0, z_0) + 2g'(t_0)\partial_1 F_0(t_0, z_0)/g(t_0) = 0$ , the rate of convergence changes because the bias is of a different asymptotic order. This is in line with results for other kernel smoothers in case of vanishing first order bias terms.

The proof of this theorem, a combination of the Lindeberg-Feller Central Limit Theorem and the Delta-method, is given in the Appendix.

To illustrate the pointwise asymptotic results we simulate  $m = 1\,000$  times a sample of size  $n = 5\,000$ , using  $F_0(x, y) = \frac{1}{2}xy(x+y)$  for  $x, y \in [0, 1]$  and  $g(t) = 2t$  for  $t \in [0, 1]$ . For each sample we determine the estimator  $\hat{F}_n^{(1)}(0.5, 0.5)$  (using kernel density  $k(y) = \frac{3}{4}(1-y^2)1_{[-1,1]}(y)$  and smoothing parameter  $\alpha_n = 0.09$ ) and the resulting value of  $n^{2/5}(\hat{F}_n^{(1)}(0.5, 0.5) - F_0(0.5, 0.5))$ . Figure 2 shows these  $m$  values, in a QQ-plot (with the line  $y = \mu_1 + x\sigma$ ) as well as in a histogram (with the  $\mathcal{N}(\mu_1, \sigma^2)$  density). Here  $\mu_1$  and  $\sigma$  are as defined in (10) and (11) for this  $F_0$  and  $g$ .

[Figure 2 here]

Under definition (K.1) and assumptions (K.2) and (K.3) on the kernel densities  $k$  and  $\tilde{k}$ , we can prove that for  $t_0, z_0 > 0$  fixed  $n^{2/5}(\hat{F}_n^{(2)}(t_0, z_0) - \hat{F}_n^{(1)}(t_0, z_0))$  converges to zero in probability whenever  $\beta_n$  converges faster to zero than  $n^{-1/5}$ . As a consequence, these estimators are (first order) asymptotically equivalent. For  $\beta_n$  tending to zero slower than  $n^{-1/5}$ ,  $n^{2/5}|\hat{F}_n^{(2)}(t_0, z_0) -$

$|\hat{F}_n^{(1)}(t_0, z_0)| \rightarrow \infty$  in probability. These results are more precisely stated in Theorem 4 and Corollary 5.

**Theorem 4** *Assume  $F_0$  and  $g$  satisfy conditions (F.1) and (G.1). Also assume  $k$  and  $\tilde{k}$  satisfy conditions (K.2) and (K.3). Fix  $t_0, z_0 > 0$  such that  $\partial_2^2 F_0(t, z)$  and  $g(t)$  exist and are continuous in a neighborhood of  $(t_0, z_0)$  and  $t_0$ , respectively, and  $\partial_2^2 F_0(t_0, z_0) \neq 0$  and  $g(t_0) \neq 0$ . Let  $\alpha_n = c_1 n^{-1/5}$  and  $\beta_n = c_2 n^{-\beta}$ , then for  $\beta > 1/5$*

$$n^{2/5}(\hat{F}_n^{(2)}(t_0, z_0) - \hat{F}_n^{(1)}(t_0, z_0)) \xrightarrow{\mathcal{P}} 0,$$

for  $\beta = 1/5$

$$n^{2/5}(\hat{F}_n^{(2)}(t_0, z_0) - \hat{F}_n^{(1)}(t_0, z_0)) \xrightarrow{\mathcal{P}} \frac{1}{2} c_2^2 m_2(k) \partial_2^2 F_0(t_0, z_0)$$

while for  $\beta < 1/5$   $n^{2/5}|\hat{F}_n^{(2)}(t_0, z_0) - \hat{F}_n^{(1)}(t_0, z_0)| \xrightarrow{\mathcal{P}} \infty$ .

The proof of this theorem is given in the Appendix.

As a consequence of this theorem, the estimators  $\hat{F}_n^{(1)}$  and  $\hat{F}_n^{(2)}$  are pointwise asymptotically equivalent for  $\beta > 1/5$ , while for  $\beta = 1/5$ ,  $\hat{F}_n^{(2)}(t_0, z_0)$  has an additional (possibly negative) asymptotic bias term.

**Corollary 5** *In addition to the conditions of Theorem 3, assume  $\partial_2^2 F_0(t_0, z_0) \neq 0$  and  $g(t_0) \neq 0$ . Let  $\alpha_n = c_1 n^{-1/5}$  and  $\beta_n = c_2 n^{-\beta}$ . Then for  $\beta > 1/5$*

$$n^{2/5}(\hat{F}_n^{(2)}(t_0, z_0) - F_0(t_0, z_0)) \rightsquigarrow \mathcal{N}(\mu_1, \sigma^2)$$

where  $\mu_1$  and  $\sigma^2$  are defined in (10) and (11) (with  $c = c_1$ ). For  $\beta = 1/5$

$$n^{2/5}(\hat{F}_n^{(2)}(t_0, z_0) - F_0(t_0, z_0)) \rightsquigarrow \mathcal{N}(\mu_2, \sigma^2),$$

where

$$\mu_2 = \mu_1 + \frac{1}{2} c_2^2 m_2(\tilde{k}) \partial_2^2 F_0(t_0, z_0). \quad (12)$$

*Proof:* This immediately follows from Theorem 4.  $\square$

Figure 3 shows the values of  $n^{2/5}(\hat{F}_n^{(2)}(0.5, 0.5) - \hat{F}_n^{(1)}(0.5, 0.5))$  as a function of  $n$  with  $\alpha_n = \frac{1}{2} n^{-1/5}$  and  $\beta_n = \frac{1}{2} n^{-1/3}$ . The solid lines are the lines  $\pm \frac{1}{2} n^{-1/6}$ , the order of the standard deviation of  $n^{2/5}(\hat{F}_n^{(2)}(0.5, 0.5) - \hat{F}_n^{(1)}(0.5, 0.5))$  (see the proof of Theorem 4 in the appendix). For  $F_0$  and  $g$  we used the same setting as in Figure 2, for  $\tilde{k}$  we used  $\tilde{k}(x, y) = k(x)k(y)$  the product kernel density with  $k(u) = \frac{3}{4}(1 - u^2)$  for  $u \in [-1, 1]$ .

[Figure 3 here]

Figure 4 shows  $m = 1\,000$  values of  $n^{2/5}(\hat{F}_n^{(2)}(0.5, 0.5) - F_0(0.5, 0.5))$  for  $n = 5\,000$ ,  $\alpha_n = \frac{1}{2}n^{-1/5}$  and  $\beta_n = \frac{1}{2}n^{-1/3}$ , in a QQ-plot (with the line  $y = \mu_1 + x\sigma$ ) as well as in a histogram (with the  $\mathcal{N}(\mu_1, \sigma^2)$  density). Here  $\mu_1$  and  $\sigma$  are as defined in (10) and (11) for  $F_0$ ,  $g$  and  $\tilde{k}$  the same as in Figure 3.

[Figure 4 here]

## 5 Smooth functionals

It is well known that in the current status model certain functionals of the model can be estimated at  $\sqrt{n}$  rate, although the pointwise estimation rate is lower, see, e.g., Groeneboom (1996). In the continuous marks model we have a similar situation and we briefly sketch how the theory of smooth functionals applies here. In the “hidden space” one would be allowed to observe the random variable  $(X, Y)$  with distribution function  $F$ , and the so-called score operator from functions on the hidden space to functions on the observation space is in this case given by

$$\begin{aligned} [L_F(a)](t, z, \delta) &= \mathbb{E} \{a(X, Y) | (T, Z, \Delta) = (t, z, \delta)\} \\ &= \frac{\delta \int_0^t a(x, z) dF_z(x)}{F_z(t)} + \frac{(1 - \delta) \int_{x=t}^\infty \int_{y=0}^\infty a(x, y) dF_y(x) dy}{1 - F(t, \infty)}, \end{aligned}$$

where  $F_z(x) = \partial_2 F(x, z) = \frac{\partial}{\partial z} F(x, z)$ . Note that the  $F_z$  correspond to the component sub-distribution functions in the model with finitely many competing risks and that  $F(t, \infty) = \int_0^\infty F_z(t) dz$ . Here  $L_F$  is a mapping from  $L_2^0(F)$  to  $L_2^0(H)$ , where  $L_2^0(F)$  denotes the space of square integrable functions  $a$  with zero expectation, i.e.

$$\mathbb{E}_F a(X, Y) = \int a(x, y) dF(x, y) = 0, \quad \mathbb{E}_F a(X, Y)^2 = \int a(x, y)^2 dF(x, y) < \infty. \quad (13)$$

Similarly,  $L_2^0(H)$  is the space of functions  $b$  with the properties:

$$\mathbb{E}_H b(T, Z, \Delta) = \int b(t, z, \delta) dH(t, z, \delta) = 0, \quad \mathbb{E}_H b(T, Z, \Delta)^2 = \int b(t, z, \delta)^2 dH(t, z, \delta) < \infty.$$

Using the first relation in (13) we get:

$$\begin{aligned} [L_F(a)](t, z, \delta) &= \frac{\delta \int_0^t a(x, z) dF_z(x)}{F_z(t)} + \frac{(1 - \delta) \int_{x=t}^\infty \int_{y=0}^\infty a(x, y) dF_y(x) dy}{1 - F(t, \infty)} \\ &= \frac{\delta \int_0^t a(x, z) dF_z(x)}{F_z(t)} - \frac{(1 - \delta) \int_{x=0}^t \int_{y=0}^\infty a(x, y) dF_y(x) dy}{1 - F(t, \infty)}. \end{aligned}$$

We now consider the adjoint of  $L_F$ , mapping the functions  $b \in L_2^0(H)$  back into  $L_2^0(F)$ . The adjoint is given by:

$$[L_F^*(b)](x, y) = \int_{t=x}^{\infty} b(t, y, 1) dG(t) + \int_{t=0}^x b(t, 0, 0) dG(t).$$

This is analogous to what we get in the current status model, see e.g., Groeneboom (1996).

In order to make this somewhat more concrete, we consider the functional

$$\mu_F = \int x dF_{0,X}(x) = \int x dF(x, \infty). \quad (14)$$

Then the score function in the hidden space is:

$$a(x, y) = x - \int x dF(x, \infty) = x - \iint u dF_w(u) dw,$$

so only depends on the first argument, and we have to solve the equation

$$\int_{t=x}^{\infty} b(t, z, 1) dG(t) + \int_{t=0}^x b(t, 0, 0) dG(t) = x - \iint u dF_w(u) dw,$$

where  $b$  has to be in the (closure of the) range of the score operator, so this would be

$$b(t, z, \delta) = \frac{\delta \int_0^t a(x, z) dF_z(x)}{F_z(t)} - \frac{(1 - \delta) \int_{x=0}^t \int_{y=0}^{\infty} a(x, y) dF_y(x) dy}{1 - F(t, \infty)}, \text{ for some } a,$$

if  $b$  is in the range itself (and not only its closure). We therefore consider the equation:

$$\int_{t=x}^{\infty} \frac{\int_{u=0}^t a(u, z) dF_z(u)}{F_z(t)} dG(t) - \int_{t=0}^x \frac{\int_{u=0}^t \int_{y=0}^{\infty} a(u, y) dF_y(u) dy}{1 - F(t, \infty)} dG(t) = x - \iint u dF_w(u) dw.$$

Differentiation w.r.t.  $x$  yields:

$$-\frac{\int_{u=0}^x a(u, z) dF_z(u)}{F_z(x)} - \frac{\int_{u=0}^x \int_{y=0}^{\infty} a(u, y) dF_y(u) dy}{1 - F(x, \infty)} = \frac{1}{g(x)}.$$

Letting  $\phi(x, z) = \int_{u=0}^x a(u, z) dF_z(u)$ , this is solved by taking

$$\phi(x, z) = -\frac{F_z(x)(1 - F(x, \infty))}{g(x)}.$$

So we get

$$\begin{aligned} b(t, z, \delta) &= -\frac{\delta F_z(t)(1 - F(t, \infty))}{F_z(t)g(t)} + \frac{(1 - \delta)(1 - F(t, \infty)) \int_{y=0}^{\infty} F_y(t) dy}{(1 - F(t, \infty))g(t)} \\ &= -\frac{\delta(1 - F(t, \infty))}{g(t)} + \frac{(1 - \delta)F(t, \infty)}{g(t)}, \end{aligned}$$

implying that the efficient asymptotic variance for estimating the mean functional  $\mu_F$ , defined by (14), is given by:

$$\int b(t, z, \delta)^2 dH(t, z, \delta) = \int \frac{F(t, \infty)(1 - F(t, \infty))}{g(t)} dt, \quad (15)$$

which (not surprisingly) is the same expression as one gets in the current status model.

The next question becomes whether taking  $\int x d\hat{F}_n(x, \infty)$ , where  $\hat{F}_n$  is one of our proposed estimators, will lead to an efficient estimate of  $\mu_F$ , in the sense that it converges at rate  $\sqrt{n}$ , with an asymptotic variance which attains the information lower bound (15).

Let us consider the estimator, defined by (4), and more specifically, the estimator obtained by taking  $k(y) = \frac{1}{2}1_{[-1,1]}(y)$ . Then (4) becomes

$$\hat{F}_n^{(1)}(x, z) = \frac{\int_{u \in [x-\alpha_n, x+\alpha_n], y \in (0, z]} d\mathbb{H}_n(u, y, 1)}{\int_{u \in [x-\alpha_n, x+\alpha_n], y \geq 0} d\mathbb{H}_n(u, y, \delta)},$$

where  $\mathbb{H}_n$  is the empirical distribution of the sample  $W_1, \dots, W_n$ . Also assume that  $f$  has compact support, say  $[0, 1]^2$ , as in the setting of Figure 2. Then we get as the estimate of  $F_{0,X}$ :

$$\hat{F}_n^{(1)}(x, 1) = \hat{F}_n^{(1)}(x, \infty) = \frac{\int_{u \in [x-\alpha_n, x+\alpha_n], y > 0} d\mathbb{H}_n(u, y, 1)}{\int_{u \in [x-\alpha_n, x+\alpha_n], y \geq 0} d\mathbb{H}_n(u, y, \delta)}.$$

To see whether this estimator leads to an efficient estimate of  $\mu_F$ , we have to perform a bias-variance analysis. We first consider the bias. Let  $F_{\alpha_n}$  be defined by

$$F_{\alpha_n}(x) = \frac{\int_{u \in [x-\alpha_n, x+\alpha_n], y > 0} dH_{F_0}(u, y, 1)}{\int_{u \in [x-\alpha_n, x+\alpha_n], y \geq 0} dH_{F_0}(u, y, \delta)},$$

where  $H_{F_0}$  is the distribution function of  $(T, Z, \Delta)$  in the observation space. Then

$$\begin{aligned} \int x dF_{\alpha_n}(x) &= \int_0^1 (1 - F_{\alpha_n}(x)) dx = \int_0^1 \frac{\int_{u \in [x-\alpha_n, x+\alpha_n]} dH_{F_0}(u, 0, 0)}{\int_{u \in [x-\alpha_n, x+\alpha_n], y \geq 0} dH_{F_0}(u, y, \delta)} dx \\ &= \int_{u=-\alpha_n}^{\alpha_n} \int_{x \in [0, u+\alpha_n]} \frac{g(u)(1 - F_0(u, 1))}{\int_{x-\alpha_n}^{x+\alpha_n} g(v) dv} dx du + \int_{u=-\alpha_n}^{1-\alpha_n} \int_{x \in [u-\alpha_n, u+\alpha_n]} \frac{g(u)(1 - F_0(u, 1))}{\int_{x-\alpha_n}^{x+\alpha_n} g(v) dv} dx du \\ &\quad + \int_{u=1-\alpha_n}^{1+\alpha_n} \int_{x \in [u-\alpha_n, 1]} \frac{g(u)(1 - F_0(u, 1))}{\int_{x-\alpha_n}^{x+\alpha_n} g(v) dv} dx du \end{aligned}$$

We have, if  $g$  is twice continuously differentiable and stays away from zero on  $[0, 1]$

$$\begin{aligned} \int_{x \in [u-\alpha_n, u+\alpha_n]} \frac{1}{G(x+\alpha_n) - G(x-\alpha_n)} dx &= \int_{x \in [u-\alpha_n, u+\alpha_n]} \frac{1}{2\alpha_n g(x) + \frac{1}{6}g''(x)\alpha_n^3 + \dots} dx \\ &= \int_{x \in [u-\alpha_n, u+\alpha_n]} \frac{1}{2\alpha_n g(x)(1 + O(\alpha_n^2))} dx = \frac{1}{g(u)} + O(\alpha_n^2), \end{aligned}$$

and hence

$$\int_{u=-\alpha_n}^{\alpha_n} \int_{x \in [0, u+\alpha_n]} \frac{g(u)(1 - F_0(u, 1))}{\int_{x-\alpha_n}^{x+\alpha_n} g(v) dv} dx du = \int_{u=0}^{\alpha_n} (1 - F_0(u, 1)) du + O(\alpha_n^2).$$

We also have

$$\int_{u=\alpha_n}^{1-\alpha_n} \int_{x \in [u-\alpha_n, u+\alpha_n]} \frac{g(u)(1 - F_0(u, 1))}{\int_{x-\alpha_n}^{x+\alpha_n} g(v) dv} dx du = \int_{u=\alpha_n}^{1-\alpha_n} (1 - F_0(u, 1)) du + O(\alpha_n^2),$$

and similarly

$$\int_{u=1-\alpha_n}^{1+\alpha_n} \int_{x \in [u-\alpha_n, 1]} \frac{g(u)(1-F_0(u, 1))}{\int_{x-\alpha_n}^{x+\alpha_n} g(v) dv} dx du = \int_{u=1-\alpha_n}^1 (1-F_0(u, 1)) du + O(\alpha_n^2).$$

So we obtain

$$\int_0^1 (1-F_{\alpha_n}(x)) dx = \int_0^1 (1-F_0(x, 1)) dx + O(\alpha_n^2). \quad (16)$$

Empirical process methods give us

$$\int (\hat{F}_n^{(1)}(x, 1) - F_{\alpha_n}(x)) dx = O_p(n^{-1/2}). \quad (17)$$

So (16) and (17) give us that, if (for example)  $\alpha_n$  is of order  $n^{-1/3}$ ,

$$\int (\hat{F}_n^{(1)}(x, 1) - F_0(x, 1)) dx = O_p(n^{-1/2}). \quad (18)$$

Note that this does not follow if  $\alpha_n$  is of order  $n^{-1/5}$ , since the bias term is too large in that case!

For the asymptotic variance, one has to analyze:

$$\int_{x=0}^1 \left\{ \frac{\int_{u \in [x-\alpha_n, x+\alpha_n]} d\mathbb{H}_n(u, 0, 0)}{\int_{u \in [x-\alpha_n, x+\alpha_n], y \in [0, 1]} d\mathbb{H}_n(u, y, \delta)} - \frac{\int_{u \in [x-\alpha_n, x+\alpha_n]} dH_{F_0}(u, 0, 0)}{\int_{u \in [x-\alpha_n, x+\alpha_n], y \in [0, 1]} dH_{F_0}(u, y, \delta)} \right\} dx,$$

which can be written as

$$\begin{aligned} & \int_{x=0}^1 \frac{\int_{u \in [x-\alpha_n, x+\alpha_n]} d(\mathbb{H}_n - H_{F_0})(u, 0, 0)}{\int_{u \in [x-\alpha_n, x+\alpha_n], y \in [0, 1]} d\mathbb{H}_n(u, y, \delta)} dx \\ & \quad - \int_{x=0}^1 \frac{\int_{u \in [x-\alpha_n, x+\alpha_n], y \in [0, 1]} d(\mathbb{H}_n - H_{F_0})(u, y, \delta)}{\int_{u \in [x-\alpha_n, x+\alpha_n], y \in [0, 1]} d\mathbb{H}_n(u, y, \delta)} \frac{\int_{u \in [x-\alpha_n, x+\alpha_n]} dH_{F_0}(u, 0, 0)}{\int_{u \in [x-\alpha_n, x+\alpha_n], y \in [0, 1]} dH_{F_0}(u, y, \delta)} dx \\ & \sim \int_{x=0}^1 \frac{F_0(x, 1) \int_{u \in [x-\alpha_n, x+\alpha_n]} d(\mathbb{H}_n - H_{F_0})(u, 0, 0)}{2g(x)\alpha_n} dx \\ & \quad - \int_{x=0}^1 \frac{(1-F_0(x, 1)) \int_{u \in [x-\alpha_n, x+\alpha_n], y \in (0, 1]} d(\mathbb{H}_n - H_{F_0})(u, y, 1)}{2g(x)\alpha_n} dx \\ & \sim \int_{u \in [0, 1]} \frac{F_0(u, 1)}{g(u)} d(\mathbb{H}_n - H_{F_0})(u, 0, 0) - \int_{u \in [0, 1], y \in (0, 1]} \frac{1-F_0(u, 1)}{g(u)} d(\mathbb{H}_n - H_{F_0})(u, y, 1). \end{aligned}$$

So the asymptotic variance is given by:

$$\begin{aligned} & \int_{u \in [0, 1]} \frac{F_0(u, 1)^2}{g(u)^2} dH_{F_0}(u, 0, 0) + \int_{u \in [0, 1], y \in (0, 1]} \frac{(1-F_0(u, 1))^2}{g(u)^2} dH_{F_0}(u, y, 1) \\ & = \int_0^1 \frac{F_0(u, 1)^2(1-F_0(u, 1))}{g(u)} du + \int_0^1 \frac{(1-F_0(u, 1))^2 F_0(u, 1)}{g(u)} du \\ & = \int_0^1 \frac{F_0(u, 1)(1-F_0(u, 1))}{g(u)} du = \int_0^1 \frac{F_0(u, \infty)(1-F_0(u, \infty))}{g(u)} du. \end{aligned}$$

The conclusion is that in this example, our estimator of  $\mu_F$  converges at rate  $\sqrt{n}$  and that its asymptotic variance attains the information lower bound, *provided the bandwidth  $\alpha_n$  tends to*

zero faster than  $n^{-1/4}$ . It also illustrates that a bandwidth of order  $n^{-1/5}$ , which is an obvious choice for the pointwise estimation, is not suitable if we want to estimate smooth functionals, a phenomenon that seems (more or less) well known. Similar analyses can be performed for other smooth functionals, but since the local estimation problem is the main focus of our paper, we will not pursue this further here.

## 6 Simulation study

The estimators  $\hat{F}_n^{(1)}$  and  $\hat{F}_n^{(2)}$  are asymptotically equivalent for sufficiently small choices of the smoothing parameter  $\beta_n$ . To get some insight in the finite sample differences between the estimators, we run a simulation study. We simulated data according to  $F_0(x, y) = \frac{1}{2}xy(x + y)$  for  $x, y \in [0, 1]$  and  $g(t) = 2t$  for  $t \in [0, 1]$  for different sample sizes  $n = 500$ ,  $n = 1000$ ,  $n = 5000$  and  $n = 10000$ . For each simulation we computed the estimators  $\hat{F}_n^{(1)}(t_0, z_0)$  and  $\hat{F}_n^{(2)}(t_0, z_0)$  for two different values of  $(t_0, z_0)$  and different values of the smoothing parameters  $\alpha_n$  and  $\beta_n$ . We repeated this  $B = 250$  times, resulting in 250 estimates  $\hat{F}_{n, \alpha_n, \beta_n}^{(i), 1}(t_0, z_0), \hat{F}_{n, \alpha_n, \beta_n}^{(i), 2}(t_0, z_0), \dots, \hat{F}_{n, \alpha_n, \beta_n}^{(i), 250}(t_0, z_0)$  ( $i = 1, 2$ ) for each value of the smoothing parameters  $\alpha_n$  and  $\beta_n$ . Then, we estimated the Mean Squared Error (MSE) of the estimator  $\hat{F}_n^{(i)}(t_0, z_0)$  by

$$\frac{1}{B} \sum_{j=1}^B \left( \hat{F}_{n, \alpha_n, \beta_n}^{(i), j}(t_0, z_0) - F_0(t_0, z_0) \right)^2.$$

Table 1 shows the minimum value of the estimated MSE for each estimator, for each  $n$  and in two different points  $(t_0, z_0)$ . It also shows the values of the smoothing parameters  $\alpha_n$  and  $\beta_n$  that yielded this value. The standard error of the mean of the squared differences  $\left( \hat{F}_{n, \alpha_n, \beta_n}^{(i), j}(t_0, z_0) - F_0(t_0, z_0) \right)^2$  are given in brackets. The binned MLE  $\tilde{F}_n$  studied by Maathuis and Wellner (2008) and the Maximum Smoothed Likelihood Estimator (MSLE)  $\hat{F}_n^{MS}$  studied by Groeneboom et al. (2010) are included in this simulation study.

[Table 1 here]

Figure 5 shows the resulting values of estimated MSEs as function of  $\alpha_n$ . For  $\tilde{F}_n$ , the smoothing parameter  $\alpha_n$  is the binwidth in  $z$ -direction, for  $\hat{F}_n^{MS}$  we have that  $\alpha_n$  and  $\beta_n$  are the binwidths in  $t$ - and  $z$ -direction, respectively. Both  $\hat{F}_n^{(2)}$  and  $\hat{F}_n^{MS}$  depend on two smoothing parameters, and we fixed the value of  $\beta_n$  to be equal to that value that yielded the overall minimal estimated MSEs of the estimators. Determining the optimal value(s) of the smoothing parameter(s) for  $\tilde{F}_n$  and  $\hat{F}_n^{MS}$  was a bit tedious; the estimated MSE of  $\tilde{F}_n$  was very wiggly, the estimated MSE of  $\hat{F}_n^{MS}$  is only nicely  $U$ -shaped for bigger values of  $n$  due

to computational issues. Although we choose the values of  $\alpha_n$  and  $\beta_n$  also as the minimizing binwidths of the estimated MSEs, these choices might not be good estimates.

[Figure 5 here]

This simulation study, of which only some results are illustrated in Figure 5 for  $\hat{F}_n^{(1)}$  and  $\hat{F}_n^{(2)}$  only, shows that the estimated MSEs of both plug-in inverse estimators are almost equal. Based on the estimated MSEs and the standard errors of the mean of the squared differences between the estimators  $\hat{F}_n^{(1)}$  and  $\hat{F}_n^{(2)}$  and the true distribution function, confidence intervals can be computed. The intervals for  $\hat{F}_n^{(1)}$  and  $\hat{F}_n^{(2)}$  have non-empty intersections, implying that for this specific example there is no significant finite sample difference between the smooth plug-in inverse estimators.

## 7 Bandwidth selection in practice

The estimators  $\hat{F}_n^{(1)}$  and  $\hat{F}_n^{(2)}$  depend on smoothing parameters  $\alpha_n$  and  $\beta_n$  (only  $\hat{F}_n^{(2)}$ ). As with usual kernel density estimators, the estimators are quite sensitive to the choice of the smoothing parameters. Small values of  $\alpha_n$  and  $\beta_n$  will result in wiggly estimators reflecting the high variance, whereas big values of  $\alpha_n$  and  $\beta_n$  will give smooth stable, but biased, estimators. One way to obtain good smoothing parameters that depend on the data is via the smoothed bootstrap.

The focus of this paper is on the pointwise asymptotic behavior of the estimators  $\hat{F}_n^{(1)}$  and  $\hat{F}_n^{(2)}$ , so also the choice of  $\alpha_n$  and  $\beta_n$  is only considered locally at the point  $(t_0, z_0)$ . The smoothed bootstrap differs from the empirical bootstrap in the distribution it samples from. In the empirical bootstrap one samples from the empirical distribution function of the data, whereas in the smoothed bootstrap one samples from a usually slightly oversmoothed estimator for the observation density  $h_{F_0}$ .

We now describe this method more specifically in our model. Let  $\hat{g}_{n,\alpha_0} =: \hat{g}_0$  and  $\hat{F}_{n,\alpha_0,\beta_0}^{(2)} =: \hat{F}_0$  be the kernel estimator and the smooth plug-in inverse estimator for  $g$  and  $F_0$ , respectively, with smoothing parameters  $\alpha_0$  and  $\beta_0$ . Then,  $(X_1^{*,1}, Y_1^{*,1}), (X_2^{*,1}, Y_2^{*,1}), \dots, (X_n^{*,1}, Y_n^{*,1})$  are sampled from  $\hat{F}_0, T_1^{*,1}, T_2^{*,1}, \dots, T_n^{*,1}$  from  $\hat{g}_0$  independently of  $(X_i^{*,1}, Y_i^{*,1})$ . The variables  $\Delta_i^{*,1}$  and  $Z_i^{*,1}$  are defined as  $1_{\{X_i^{*,1} \leq T_i^{*,1}\}}$  and  $Y_i^{*,1} \cdot \Delta_i^{*,1}$ , respectively. The estimators  $\hat{F}_{n,\alpha_n,1}^{(1)}$  and  $\hat{F}_{n,\alpha_n,\beta_n,1}^{(2)}$  are determined at the point  $(t_0, z_0)$  for several values of  $\alpha_n$  and  $\beta_n$  based on the sample  $(T_1^{*,1}, Z_1^{*,1}, \Delta_1^{*,1}), \dots, (T_n^{*,1}, Z_n^{*,1}, \Delta_n^{*,1})$ . Note that now we make the dependence of the estimators on  $\alpha_n$  and  $\beta_n$  explicit in the notation of the estimators. Actually, we only need the precise values for those observations  $(T_i^{*,1}, Z_i^{*,1}, \Delta_i^{*,1})$  that fall in  $[t_0 - \alpha_n, t_0 + \alpha_n] \times [0, z_0 +$



$\beta_n] \times \{0, 1\}$ , the precise values of  $T_i^{*,1}$  for those observations that fall in  $[t_0 - \alpha_n, t_0 + \alpha_n] \times (z_0 + \beta_n, \infty] \times \{1\}$  and the numbers of observations in the various regions outside these areas (rather than their exact locations) to compute  $\hat{F}_{n,\alpha_n,1}^{(1)}(t_0, z_0)$  and  $\hat{F}_{n,\alpha_n,\beta_n,1}^{(2)}(t_0, z_0)$ . Hence, only on this strip monotonicity of  $\hat{F}_0$  is needed as well as positivity of  $\hat{F}_{n,\alpha_n,1}^{(1)}(\infty, \infty) - \hat{F}_{n,\alpha_n,1}^{(1)}(t_0 + \alpha_n, \infty)$  and  $\hat{F}_{n,\alpha_n,\beta_n,1}^{(2)}(\infty, \infty) - \hat{F}_{n,\alpha_n,\beta_n,1}^{(2)}(t_0 + \alpha_n, \infty)$ .

The procedure described above is repeated  $B$  times resulting in  $B$  estimators  $\hat{F}_{n,\alpha_n,1}^{(1)}, \dots, \hat{F}_{n,\alpha_n,B}^{(1)}$  and  $\hat{F}_{n,\alpha_n,\beta_n,1}^{(2)}, \dots, \hat{F}_{n,\alpha_n,\beta_n,B}^{(2)}$ . Then, the MSEs of  $\hat{F}_{n,\alpha_n}^{(1)}(t_0, z_0)$  and  $\hat{F}_{n,\alpha_n,\beta_n}^{(2)}(t_0, z_0)$  can be estimated by

$$\begin{aligned}\widehat{MSE}^{(1)}(\alpha_n; t_0, z_0) &= \frac{1}{B} \sum_{b=1}^B \left( \hat{F}_{n,\alpha_n,b}^{(1)}(t_0, z_0) - \hat{F}_0(t_0, z_0) \right)^2, \\ \widehat{MSE}^{(2)}(\alpha_n, \beta_n; t_0, z_0) &= \frac{1}{B} \sum_{b=1}^B \left( \hat{F}_{n,\alpha_n,\beta_n,b}^{(2)}(t_0, z_0) - \hat{F}_0(t_0, z_0) \right)^2.\end{aligned}$$

Then, choose those values of  $\alpha_n$  and  $\beta_n$  that minimize  $\widehat{MSE}^{(1)}(\alpha_n; t_0, z_0)$  and  $\widehat{MSE}^{(2)}(\alpha_n, \beta_n; t_0, z_0)$  as smoothing parameters for the estimators  $\hat{F}_n^{(1)}(t_0, z_0)$  and  $\hat{F}_n^{(2)}(t_0, z_0)$ , respectively.

Figure 6 shows the estimated MSEs for a small simulation study. In this study, we took  $n = 100$ ,  $B = 500$ ,  $\alpha_0 = \beta_0 = 0.4$ ,  $t_0 = z_0 = 0.5$  and  $F_0$  and  $g$  as in Section 6. It also shows

$$\begin{aligned}\widetilde{MSE}^{(1)}(\alpha_n; t_0, z_0) &= \frac{1}{B} \sum_{b=1}^B \left( \hat{F}_{n,\alpha_n,b}^{(1)}(t_0, z_0) - F_0(t_0, z_0) \right)^2, \\ \widetilde{MSE}^{(2)}(\alpha_n, \beta_n; t_0, z_0) &= \frac{1}{B} \sum_{b=1}^B \left( \hat{F}_{n,\alpha_n,\beta_n,b}^{(2)}(t_0, z_0) - F_0(t_0, z_0) \right)^2,\end{aligned}$$

as function of  $\alpha_n$ . For  $\widetilde{MSE}^{(2)}$  and  $\widetilde{MSE}^{(2)}$  it only shows the estimates for that value of  $\beta_n$  that has the smallest estimated MSE.

**[Figure 6 here]**

There are other methods to obtain data-dependent bandwidths, for example via cross-validation (Rudemo 1982). Usually in cross-validation methods a global risk measure is minimized (like the Integrated MSE), hence its minimizer can be used as a global optimal bandwidth.

## 8 Concluding remarks

In this paper we consider two plug-in inverse estimators for the distribution function of the vector  $(X, Y)$  in the current status continuous mark model. The first estimator  $\hat{F}_n^{(1)}$  is shown to be consistent and pointwise asymptotically normally distributed. However,  $\hat{F}_n^{(1)}$  does not have a Lebesgue density, since it only puts mass on the lines  $[0, \infty) \times \{Z_i\}$  with  $Z_i > 0$  for  $i = 1, \dots, n$ .

The second estimator,  $\hat{F}_n^{(2)}$ , does have a Lebesgue density. For a range of possible choices of the bandwidths  $\alpha_n$  and  $\beta_n$  we establish consistency of this estimator. Taking  $\alpha_n = n^{-1/5}$  and  $\beta_n = n^{-\beta}$ , we prove that asymptotically for  $\beta < 3/10$  the Lebesgue density of  $\hat{F}_n^{(2)}$  is positive on a region where  $f_0$  is positive which stays away from the boundary of its support. This means that, although for finite sample size  $n$  the estimator  $\hat{F}_n^{(2)}$  need not be a bivariate distribution function, “isotonisation” of it is not necessary asymptotically. Put differently, any common shape regularized version of our estimator is asymptotically equivalent with our estimator. However, this only holds asymptotically, and for finite sample size  $n$  it might be desirable to have an estimator which is a true bivariate distribution function, satisfying condition (6). For example when one wants to sample in a smoothed bootstrap procedure. Furthermore, we prove that  $\hat{F}_n^{(2)}$  is asymptotically normally distributed for  $\beta \geq 1/5$ . Hence, for  $\beta \in [1/5, 3/10)$ , the estimator  $\hat{F}_n^{(2)}$  asymptotically behaves as a distribution function with pointwise normal limiting distribution on a large subset of  $[0, \infty)^2$ .

## Acknowledgements

We would like to thank two anonymous referees for their valuable comments concerning the practical use of our results and the readability of the paper.

## References

- Burke, M. D. (1988), Estimation of a bivariate distribution function under random censorship, *Biometrika*, **75**: 379–382.
- Cacoullos, T. (1964), Estimation of a multivariate density, *Ann. Inst. Statist. Math.*, **18**: 179–189.
- Flynn, N. M., Forthal, D. N., and The rgp120 HIV Vaccine Study Group (2005), Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection, *Journal of Infectious Diseases*, **191**: 654–665.
- Gilbert, P., Self, S., Rao, M., Naficy, A., and Clemens, J. (2001), Sieve analysis: methods for assessing from vaccine trial data how vaccine efficacy varies with genotypic and phenotypic pathogen variation, *Journal of Clinical Epidemiology*, **54**: 68–85.
- Groeneboom, P. (1996), Lectures on inverse problems, in: *Lectures on Probability Theory and Statistics, Lecture Notes in Mathematics*, volume 1648, 67–164, Springer, Berlin.
- Groeneboom, P., Jongbloed, G., and Witte, B. I. (2010), Maximum smoothed likelihood

- estimation and smoothed maximum likelihood estimation in the current status model, *Ann. Statist.*, **38**: 352–387.
- Groeneboom, P., Maathuis, M. H., and Wellner, J. A. (2008a), Current status data with competing risks: consistency and rates of convergence of the MLE, *Ann. Statist.*, **36**: 1031–1063.
- Groeneboom, P., Maathuis, M. H., and Wellner, J. A. (2008b), Current status data with competing risks: limiting distribution of the MLE, *Ann. Statist.*, **36**: 1064–1089.
- Hall, P. and Smith, R. L. (1988), The kernel method for unfolding sphere size distributions, *J. Comput. Phys.*, **74**: 409 – 421.
- Härdle, W., Janssen, P., and Serfling, R. (1988), Strong uniform consistency rates for estimators of conditional functionals, *Ann. Statist.*, **88**: 1428–1449.
- Hudgens, M. G., Maathuis, M. H., and Gilbert, P. B. (2007), Nonparametric estimation of the joint distribution of a survival time subject to interval censoring and a continuous mark variable, *Biometrics*, **63**: 372–380.
- Maathuis, M. H. and Wellner, J. A. (2008), Inconsistency of the MLE for the joint distribution of interval censored survival times and continuous marks, *Scand. J. Statist.*, **35**: 83–103.
- Marron, J. S. and Padgett, W. J. (1987), Asymptotically optimal bandwidth selection for kernel density estimators from randomly right-censored samples, *Ann. Statist.*, **15**: 1520–1535.
- Mokkadem, A., Pelletier, M., and Worms, J. (2005), Large and moderate deviations principles for kernel estimation of a multivariate density and its partial derivatives, *Aust. N. Z. J. Stat.*, **47**: 489–502.
- Patil, P. N., Wells, M. T., and Marron, J. S. (1994), Some heuristics of kernel based estimators of ratio functions, *J. Nonparametr. Stat.*, **4**: 203–209.
- Rudemo, M. (1982), Empirical choice of histograms and kernel density estimators, *Scandinavian Journal of Statistics*, **9**: 65–78.
- Silverman, B. W. (1978), Weak and strong uniform consistency of the kernel estimate of a density and its derivative, *Ann. Statist.*, **6**: 177–184.
- Stefanski, L. and Carroll, R. J. (1990), Deconvoluting kernel density estimators, *Statistics*, **21**: 169–184.

B.I. Witte, VU University Medical Center, Department of Epidemiology and Biostatistics, PO Box 7057, 1007 BM Amsterdam, The Netherlands  
E-mail: B.Witte@vumc.nl

## A Technical lemmas and proofs

**Lemma 6** *Assume that  $F_0$  and  $g$  satisfy conditions (F.1) and (G.1). Let  $\mathcal{S}$  and  $\mathcal{M}_\delta$  be as defined in Theorem 2. Then  $\text{proj}_X \mathcal{M}_\delta = \{t : (t, z) \in \mathcal{M}_\delta \text{ for some } z\}$  is a closed subset of  $\mathcal{S}_{0,X}^\circ$ .*

*Proof:* Fix  $\delta > 0$ . To prove that  $\text{proj}_X \mathcal{M}_\delta$  is a closed subset of  $\mathcal{S}_{0,X}^\circ$  we prove that

- (i)  $\mathcal{M}_\delta$  is closed in  $[0, \infty)^2$ ,
- (ii)  $\text{proj}_X \mathcal{M}_\delta$  is closed in  $[0, \infty)$ ,
- (iii)  $\text{proj}_X \mathcal{M}_\delta$  is a subset of  $\mathcal{S}_{0,X}^\circ$ .

We now start with proving (i). By definition of  $\mathcal{S}$ , there exists an open set  $\mathcal{U} \supset \mathcal{S}$  on which  $f_0$  is uniformly continuous. Define

$$\mathcal{U}_\delta = \{(t, z) \in \mathcal{U} : f_0(t, z) \geq 2\delta\} = f_0^{-1}[[2\delta, \infty)].$$

The function  $f_0$  is continuous on  $\mathcal{U}$ , hence  $\mathcal{U}_\delta$  is closed. Since we also have that

$$\mathcal{M}_\delta = \{(t, z) \in [0, \infty)^2 : f_0(t, z) \geq 2\delta\} \cap \mathcal{S} = \mathcal{U}_\delta \cap \mathcal{S},$$

$\mathcal{M}_\delta$  is the intersection of two closed sets, hence closed itself.

For proving (ii), assume  $\text{proj}_X \mathcal{M}_\delta$  is not closed. Then, there exists a sequence  $(x_n)_n \in \text{proj}_X \mathcal{M}_\delta$  with  $x_n \rightarrow x \notin \text{proj}_X \mathcal{M}_\delta$ . By (i), the set  $\mathcal{M}_\delta$  is closed, hence by definition of  $\text{proj}_X \mathcal{M}_\delta$  there exists a sequence  $(x_n, y_n)_n \in \mathcal{M}_\delta$ . By compactness of  $\mathcal{M}_\delta$  (this follows from (i)), there exists a subsequence  $(n_k)_k$  and  $(x, y) \in \mathcal{M}_\delta$  such that  $(x_{n_k}, y_{n_k})_k \rightarrow (x, y)$ . From this it follows that  $x \in \text{proj}_X \mathcal{M}_\delta$ . This yields a contradiction, hence  $\text{proj}_X \mathcal{M}_\delta$  is closed.

To prove (iii), first note that by uniform continuity of  $f_0$  on  $\mathcal{M}_\delta$

$$\exists \eta > 0 \text{ such that } \forall (t, z), (s, y) \in \mathcal{M}_\delta : \|(t, z) - (s, y)\| < \eta \implies |f_0(t, z) - f_0(s, y)| < \delta.$$

Now take  $t \in \text{proj}_X \mathcal{M}_\delta$ . Then for all  $s$  in a small neighborhood of  $t$  and  $z_s > 0$  such that  $(s, z_s) \in \mathcal{M}_\delta$

$$f_{0,X}(s) = \int_0^\infty f_0(s, z) dz \geq \int_{[z_s, z_s + \eta/2]} f_0(s, z) dz \geq \delta\eta/2,$$

hence  $t \in \mathcal{S}_{0,X}^\circ$ . □

**Lemma 7** Assume that  $F_0$  and  $g$  satisfy conditions (F.1) and (G.1). Let  $\mathcal{S}$  and  $\mathcal{M}_\delta$  be as defined in Theorem 2. Then

$$\exists l_g > 0, u_g < \infty : l_g \leq g(t) \leq u_g \text{ for all } t \in \text{proj}_X \mathcal{M}_\delta. \quad (19)$$

*Proof:* The set  $\text{proj}_X \mathcal{M}_\delta$  is a closed subset of  $\mathcal{S}_{0,X}^\circ$  by Lemma 6. On  $\mathcal{S}_{0,X}^\circ$ , we have that  $0 < g < \infty$ , hence also on  $\text{proj}_X \mathcal{M}_\delta$ . Now assume (19) does not hold. Then, there exists a sequence  $(t_n)_n \rightarrow t \in \text{proj}_X \mathcal{M}_\delta$  such that  $g(t_n) \rightarrow 0$ . By the uniform continuity of  $g$ , this implies that  $g(t) = 0$ , yielding a contradiction.

The existence of  $u_g$  follows immediately from (G.1).  $\square$

**Lemma 8** Assume  $F_0$  and  $g$  satisfy conditions (F.1) and (G.1). Also assume  $k$  and  $\tilde{k}$  satisfy conditions (K.2) and (K.3). In addition, assume  $k'$  and  $\partial_1 \tilde{k}$  are uniformly continuous. Furthermore, let  $\alpha_n$  and  $\beta_n$  satisfy condition (C.2). Let  $\mathcal{S} \subset [0, \infty)^2$  be compact and such that  $f_0$  is uniformly continuous on an open subset that contains  $\mathcal{S}$  and define  $\|f\|_{\mathcal{S},\infty} = \sup_{(x,y) \in \mathcal{S}} |f(x,y)|$  and  $\mathcal{S}_X = \text{proj}_X \mathcal{S}$ . Then

$$\|\hat{g}_n - g\|_{\mathcal{S}_X,\infty} \xrightarrow{\mathcal{P}} 0, \quad \|\hat{g}'_n - g'\|_{\mathcal{S}_X,\infty} \xrightarrow{\mathcal{P}} 0 \quad (20)$$

$$\|\hat{h}_{n,1}^{(2)} - h_1\|_{\mathcal{S},\infty} \xrightarrow{\mathcal{P}} 0, \quad \|\partial_1 \hat{h}_{n,1}^{(2)} - \partial_1 h_1\|_{\mathcal{S},\infty} \xrightarrow{\mathcal{P}} 0 \quad (21)$$

*Proof:* The results in (20) follow directly from Theorem A and C in Silverman (1978). The first result in (21) follows from Theorem 3.3 in Cacoullos (1964). By Theorem 3 in Mokkadem et al. (2005),

$$\lim_{n \rightarrow \infty} (n\alpha_n^3\beta_n)^{-1} \log P(\|\partial_1 \hat{h}_n - \partial_1 h_{f_0}\|_{\mathcal{S},\infty} \geq \delta) = -c,$$

for some constant  $c > 0$  only depending on  $\delta$  and  $\partial_1 h_{f_0}$ . Hence, for  $n$  sufficiently large

$$P(\|\partial_1 \hat{h}_n - \partial_1 h_{f_0}\|_{\mathcal{S},\infty} \geq \delta) \leq 2e^{-n\alpha_n^3\beta_n c} \rightarrow 0.$$

The results in Cacoullos (1964) and Mokkadem et al. (2005) hold for density estimators, whereas the estimator  $\hat{h}_{n,1}^{(2)}$  is a sub-density. However, the results in (21) follow from these after defining a binomially distributed sample size  $N_1$  and reason similarly as the proof of (A.3) in Groeneboom et al. (2010).  $\square$

*Proof of Theorem 3:* Define

$$Y_i = \begin{pmatrix} Y_{i;1} \\ Y_{i;2} \end{pmatrix} = n^{-3/5} \begin{pmatrix} k_{\alpha_n}(t_0 - T_i) \\ 1_{[0,z_0]}(Z_i) \Delta_i k_{\alpha_n}(t_0 - T_i) \end{pmatrix}.$$

By the assumptions on  $F_0$  and  $g$  and condition (K.2), we have

$$\begin{aligned} \mathbb{E} Y_{i;1} &= n^{-3/5} g(t_0) + \frac{1}{2} n^{-1} c^2 m_2(k) g''(t_0) + o(n^{-1}), \\ \text{Var } Y_{i;1} &= n^{-1} c^{-1} g(t_0) \int k^2(y) dy + O(n^{-6/5}), \\ \mathbb{E} Y_{i;2} &= n^{-3/5} F_0(t_0, z_0) g(t_0) + \frac{1}{2} n^{-1} c^2 m_2(k) \partial_1^2 \{F_0(t_0, z_0) g(t_0)\} + o(n^{-1}), \\ \text{Var } Y_{i;2} &= n^{-1} c^{-1} F_0(t_0, z_0) g(t_0) \int k^2(y) dy + O(n^{-6/5}). \end{aligned}$$

Furthermore we have

$$\text{Cov}(Y_{i;1}, Y_{i;2}) = n^{-1} c^{-1} F_0(t_0, z_0) g(t_0) \int k^2(y) dy + O(n^{-6/5}),$$

so that

$$\begin{aligned} \sum_{i=1}^n \mathbb{E} Y_i &= n^{2/5} \begin{pmatrix} g(t_0) \\ F_0(t_0, z_0) g(t_0) \end{pmatrix} + \begin{pmatrix} \frac{1}{2} c^2 m_2(k) g''(t_0) \\ \frac{1}{2} c^2 m_2(k) \partial_1^2 \{F_0(t_0, z_0) g(t_0)\} \end{pmatrix} + o(1) \\ \sum_{i=1}^n \text{Var } Y_i &= c^{-1} g(t_0) \int k(u)^2 du \begin{pmatrix} 1 & F_0(t_0, z_0) \\ F_0(t_0, z_0) & F_0(t_0, z_0) \end{pmatrix} + O(n^{-1/5}) = \Sigma_1 + O(n^{-1/5}). \end{aligned}$$

Here we denote by  $\text{Var } Y_i$  the covariance matrix of the vector  $Y_i$ . By the Lindeberg–Feller central limit theorem we then get

$$\begin{aligned} n^{2/5} \left( \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n k_{\alpha_n}(t_0 - T_i) \\ \frac{1}{n} \sum_{i=1}^n 1_{[0, z_0]}(Z_i) \Delta_i k_{\alpha_n}(t_0 - T_i) \end{pmatrix} - \begin{pmatrix} g(t_0) \\ F_0(t_0, z_0) g(t_0) \end{pmatrix} \right) \\ - \frac{1}{2} c^2 m_2(k) \begin{pmatrix} g''(t_0) \\ \partial_1^2 \{F_0(t_0, z_0) g(t_0)\} \end{pmatrix} = \sum_{i=1}^n (Y_i - \mathbb{E} Y_i) + o(1) \rightsquigarrow \mathcal{N}(0, \Sigma_1). \quad (22) \end{aligned}$$

For the pointwise asymptotic result of  $\hat{F}_n^{(1)}$ , note that

$$\hat{F}_n^{(1)}(t_0, z_0) = \phi \left( \frac{1}{n} \sum_{i=1}^n k_{\alpha_n}(t_0 - T_i), \frac{1}{n} \sum_{i=1}^n 1_{[0, z_0]}(Z_i) \Delta_i k_{\alpha_n}(t_0 - T_i) \right),$$

for  $\phi(u, v) = v/u$ . Now applying the Delta-method to (22) gives

$$n^{2/5} (\hat{F}_n^{(1)}(t_0, z_0) - F_0(t_0, z_0)) \rightsquigarrow \mathcal{N}(\mu_1, \sigma^2)$$

where  $\mu_1$  and  $\sigma^2$  are defined in (10) and (11). □

*Proof of Theorem 4:* For  $i = 1, 2$ , let  $N_n^{(i)}(t_0, z_0)$  be the numerator in the definitions (4) and (3) of  $\hat{F}_n^{(i)}(t_0, z_0)$  at a fixed point  $(t_0, z_0)$ , and note that we can write

$$\begin{aligned}
N_n^{(2)}(t_0, z_0) &= \frac{1}{n} \sum_i 1_{[0, z_0 - \beta_n]}(Z_i) \Delta_i \int_0^{z_0} \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz \\
&\quad + \frac{1}{n} \sum_i 1_{(z_0 - \beta_n, z_0]}(Z_i) \Delta_i \int_0^{z_0} \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz \\
&\quad + \frac{1}{n} \sum_i 1_{(z_0, z_0 + \beta_n]}(Z_i) \Delta_i \int_0^{z_0} \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz \\
&= \frac{1}{n} \sum_i 1_{[0, z_0 - \beta_n]}(Z_i) \Delta_i \int_0^{z_0} \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz \\
&\quad + \frac{1}{n} \sum_i 1_{(z_0 - \beta_n, z_0]}(Z_i) \Delta_i \int_0^{Z_i + \beta_n} \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz \\
&\quad - \frac{1}{n} \sum_i 1_{(z_0 - \beta_n, z_0]}(Z_i) \Delta_i \int_{z_0}^{Z_i + \beta_n} \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz \\
&\quad + \frac{1}{n} \sum_i 1_{(z_0, z_0 + \beta_n]}(Z_i) \Delta_i \int_0^{z_0} \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz \\
&= \frac{1}{n} \sum_i 1_{[0, z_0]}(Z_i) \Delta_i k_{\alpha_n}(t_0 - T_i, z_0) - \frac{1}{n} \sum_i 1_{(z_0 - \beta_n, z_0]}(Z_i) \Delta_i \int_{z_0}^{Z_i + \beta_n} \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz \\
&\quad + \frac{1}{n} \sum_i 1_{(z_0, z_0 + \beta_n]}(Z_i) \Delta_i \int_{Z_i - \beta_n}^{z_0} \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz.
\end{aligned}$$

In the last equality we use (K.1), so that

$$\begin{aligned}
R_n &= N_n^{(2)}(t_0, z_0) - N_n^{(1)}(t_0, z_0) = \frac{1}{n} \sum_{i=1}^n 1_{(z_0, z_0 + \beta_n]}(Z_i) \Delta_i \int_{Z_i - \beta_n}^{z_0} \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz \\
&\quad - \frac{1}{n} \sum_{i=1}^n 1_{(z_0 - \beta_n, z_0]}(Z_i) \Delta_i \int_{z_0}^{Z_i + \beta_n} \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz =: \frac{1}{n} \sum_{i=1}^n U_i.
\end{aligned}$$

First we consider the variance of  $n^{2/5} R_n$ . Observe that

$$\begin{aligned}
|U_i| &\leq 1_{(z_0, z_0 + \beta_n]}(Z_i) \Delta_i \int_{Z_i - \beta_n}^{z_0} \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz \\
&\quad + 1_{(z_0 - \beta_n, z_0]}(Z_i) \Delta_i \int_{z_0}^{Z_i + \beta_n} \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz \\
&\leq 1_{(z_0, z_0 + \beta_n]}(Z_i) \Delta_i \int_{Z_i - \beta_n}^{Z_i + \beta_n} \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz \\
&\quad + 1_{(z_0 - \beta_n, z_0]}(Z_i) \Delta_i \int_{Z_i - \beta_n}^{Z_i + \beta_n} \tilde{k}_{\alpha_n, \beta_n}(t_0 - T_i, z - Z_i) dz \\
&= 1_{(z_0 - \beta_n, z_0 + \beta_n]}(Z_i) \Delta_i k_{\alpha_n}(t_0 - T_i) := S_i,
\end{aligned}$$

with

$$\mathbb{E} S_i^2 = \int_u \int_v 1_{(z_0 - \beta_n, z_0 + \beta_n]}(v) k_{\alpha_n}^2(t_0 - u) h_1(u, v) dv du = \alpha_n^{-1} \beta_n 2 h_1(t_0, z_0) \int k^2(x) dx + \mathcal{O}(\beta_n).$$

Since

$$\text{Var } R_n = \frac{1}{n} \text{Var } U_1 = \frac{1}{n} \{E U_1^2 - (E U_1)^2\} \leq \frac{1}{n} E S_1^2 = O(n^{-1} \alpha_n^{-1} \beta_n),$$

$\text{Var } n^{2/5} R_n \rightarrow 0$  for  $\alpha_n = n^{-1/5}$  and  $\beta_n = n^{-\beta}$  with  $\beta > 0$ .

Now we consider the expectation of  $U_i$ .

$$\begin{aligned} E U_i &= \int_v \int_u 1_{(z_0, z_0 + \beta_n]}(v) \int_{v - \beta_n}^{z_0} \tilde{k}_{\alpha_n, \beta_n}(t_0 - u, z - v) dz h_1(u, v) du dv \\ &\quad - \int_v \int_u 1_{(z_0 - \beta_n, z_0]}(v) \int_{z_0}^{v + \beta_n} \tilde{k}_{\alpha_n, \beta_n}(t_0 - u, z - v) dz h_1(u, v) du dv \\ &= \beta_n \int_{y=-1}^0 \int_{x=-1}^1 \int_{w=-1}^y \tilde{k}(x, w) dw h_1(t_0 - \alpha_n x, z_0 - \beta_n y) dx dy \\ &\quad - \beta_n \int_{y=0}^1 \int_{x=-1}^1 \int_{w=y}^1 \tilde{k}(x, w) dw h_1(t_0 - \alpha_n x, z_0 - \beta_n y) dx dy \\ &= \beta_n h_1(t_0, z_0) \left\{ \int_{x=-1}^1 \int_{y=-1}^0 \int_{w=-1}^y \tilde{k}(x, w) dw dy dx - \int_{x=-1}^1 \int_{y=0}^1 \int_{w=y}^1 \tilde{k}(x, w) dw dy dx \right\} \\ &\quad - \beta_n \alpha_n \partial_1 h_1(t_0, z_0) \left\{ \int_{x=-1}^1 \int_{y=-1}^0 \int_{w=-1}^y x \tilde{k}(x, w) dw dy dx - \int_{x=-1}^1 \int_{y=0}^1 \int_{w=y}^1 x \tilde{k}(x, w) dw dy dx \right\} \\ &\quad - \beta_n^2 \partial_2 h_1(t_0, z_0) \left\{ \int_{x=-1}^1 \int_{y=-1}^0 \int_{w=-1}^y y \tilde{k}(x, w) dw dy dx - \int_{x=-1}^1 \int_{y=0}^1 \int_{w=y}^1 y \tilde{k}(x, w) dw dy dx \right\} \\ &\quad + O(\beta_n \alpha_n^2) + O(\beta_n^2 \alpha_n) + O(\beta_n^3) \\ &= -\beta_n h_1(t_0, z_0) \int_{x=-1}^1 \int_{w=-1}^1 w \tilde{k}(x, w) dw dx + \beta_n \alpha_n \partial_1 h_1(t_0, z_0) \int_{x=-1}^1 \int_{w=-1}^1 x w \tilde{k}(x, w) dw dx \\ &\quad + \frac{1}{2} \beta_n^2 \partial_2 h_1(t_0, z_0) \int_{x=-1}^1 \int_{w=-1}^1 w^2 \tilde{k}(x, w) dw dx + O(\beta_n \alpha_n^2) + O(\beta_n^2 \alpha_n) + O(\beta_n^3) \end{aligned}$$

where the last equality follows from changing the order of integration. By condition (K.3), the first two integrals are zero and the last integral equals  $m_2(\tilde{k})$ , so that

$$E n^{2/5} R_n \rightarrow \begin{cases} \pm \infty & \text{for } \beta < 1/5, \\ \frac{1}{2} c_2^2 m_2(\tilde{k}) g(t_0) \partial_2^2 F_0(t_0, z_0) & \text{for } \beta = 1/5, \\ 0 & \text{for } \beta > 1/5. \end{cases}$$

Applying Slutsky's Lemma to

$$n^{2/5} (\hat{F}_n^{(2)}(t_0, z_0) - \hat{F}_n^{(1)}(t_0, z_0)) = \frac{n^{2/5} R_n}{\hat{g}_n(t_0)}$$

gives the result.  $\square$



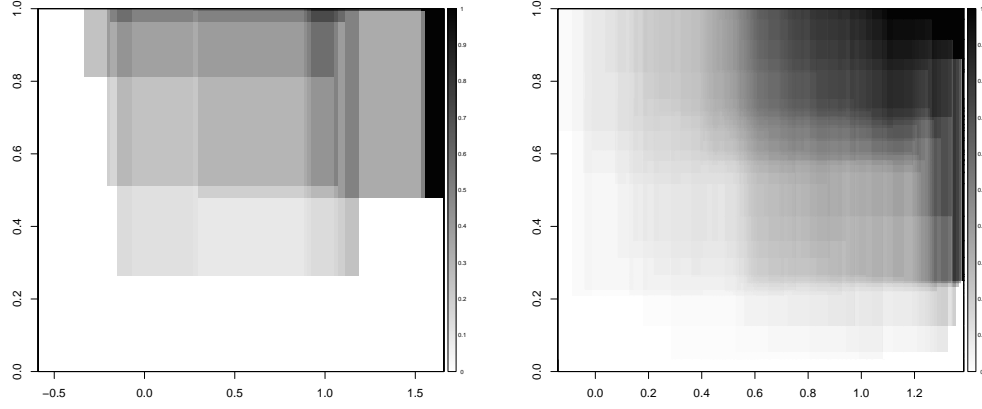


Figure 1: Two examples of the estimator  $\hat{F}_n^{(1)}$  for a sample of size  $n = 10$  (left panel) and of size  $n = 100$  (right panel),  $k(x) = \frac{1}{2}1_{[-1,1]}(x)$ ,  $\alpha_n = 0.65$  (left panel) and  $\alpha_n = 0.40$  (right panel),  $F_0(x, y) = xy$  on  $[0, 1]^2$  and  $g(t) = 1_{[0,1]}(t)$ .

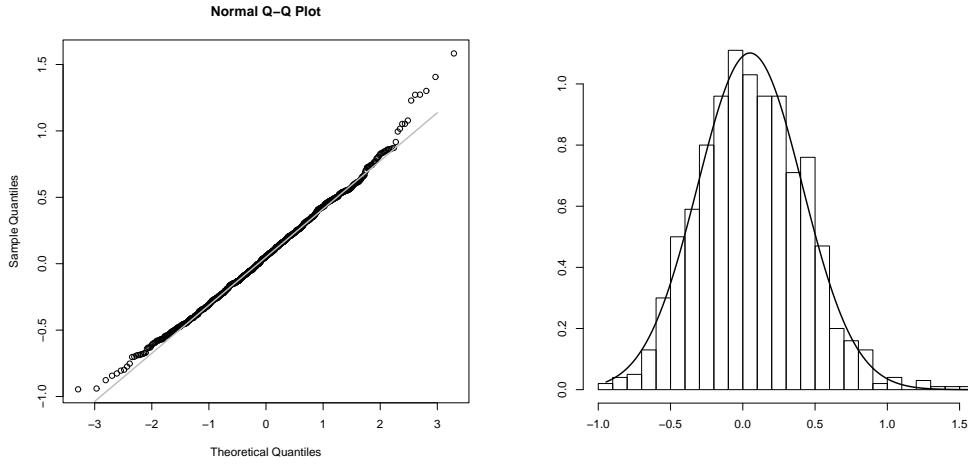


Figure 2: QQ-plot (left panel) and histogram (right panel) of  $m = 1000$  values  $n^{2/5}(\hat{F}_n^{(1)}(0.5, 0.5) - F_0(0.5, 0.5))$  for  $n = 5000$ ,  $k(y) = \frac{3}{4}(1 - y^2)1_{[-1,1]}(y)$ ,  $\alpha_n = 0.09$ ,  $F_0(x, y) = \frac{1}{2}xy(x + y)$  and  $g(t) = 2t$ , illustrating Theorem 3.

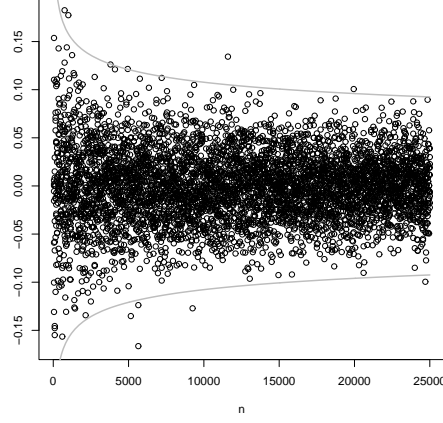


Figure 3: Values of  $n^{2/5}(\hat{F}_n^{(2)}(0.5, 0.5) - \hat{F}_n^{(1)}(0.5, 0.5))$  as a function of  $n$  for  $\tilde{k}(x, y) = k(x)k(y)$ ,  $k(u) = \frac{3}{4}(1 - u^2)$ ,  $\alpha_n = \frac{1}{2}n^{-1/5}$ ,  $\beta_n = \frac{1}{2}n^{-1/3}$ ,  $F_0(x, y) = \frac{1}{2}xy(x + y)$  and  $g(t) = 2t$ .

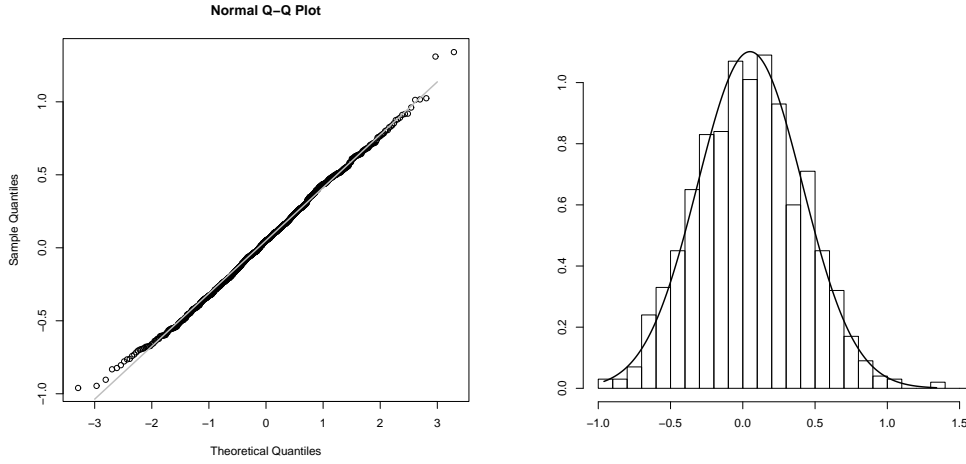
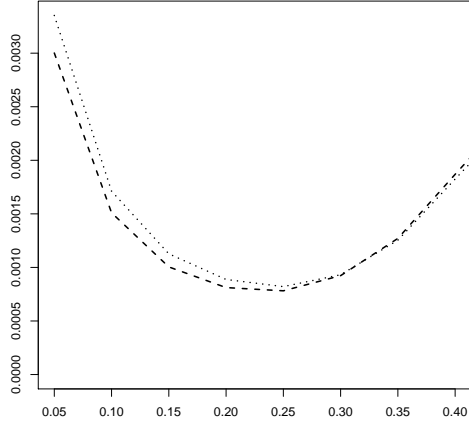


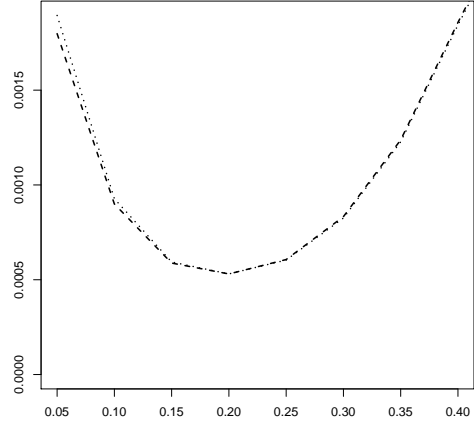
Figure 4: QQ-plot (left panel) and histogram (right panel) of  $m = 1000$  values  $n^{2/5}(\hat{F}_n^{(2)}(0.5, 0.5) - F_0(0.5, 0.5))$  for  $n = 5000$ ,  $\tilde{k}(x, y) = k(x)k(y)$ ,  $k(u) = \frac{3}{4}(1 - u^2)$ ,  $\alpha_n = 0.091$ ,  $\beta_n = 0.029$ ,  $F_0(x, y) = \frac{1}{2}xy(x + y)$  and  $g(t) = 2t$ , illustrating Corollary 5.

$(t_0, z_0)$	$n$	$\alpha_n$	$\widehat{MSE}$ (s.e.)	$(\alpha_n, \beta_n)$	$\widehat{MSE}$ (s.e.)
			$\hat{F}_n^{(1)}$		$\hat{F}_n^{(2)}$
(0.4,0.4)	500	0.20	$5.14 \cdot 10^{-4}$ ( $5.10 \cdot 10^{-5}$ )	(0.20,0.25)	$4.43 \cdot 10^{-4}$ ( $4.33 \cdot 10^{-5}$ )
	1 000	0.20	$3.31 \cdot 10^{-4}$ ( $3.04 \cdot 10^{-5}$ )	(0.20,0.15)	$3.10 \cdot 10^{-4}$ ( $3.06 \cdot 10^{-5}$ )
	5 000	0.15	$8.09 \cdot 10^{-5}$ ( $8.47 \cdot 10^{-6}$ )	(0.15,0.10)	$7.74 \cdot 10^{-5}$ ( $8.28 \cdot 10^{-6}$ )
	10 000	0.15	$4.50 \cdot 10^{-5}$ ( $3.43 \cdot 10^{-6}$ )	(0.15,0.05)	$4.50 \cdot 10^{-5}$ ( $3.38 \cdot 10^{-6}$ )
(0.6,0.6)	500	0.25	$8.21 \cdot 10^{-4}$ ( $7.48 \cdot 10^{-5}$ )	(0.25,0.15)	$7.82 \cdot 10^{-4}$ ( $7.04 \cdot 10^{-5}$ )
	1 000	0.20	$5.31 \cdot 10^{-4}$ ( $4.34 \cdot 10^{-5}$ )	(0.20,0.05)	$5.31 \cdot 10^{-4}$ ( $4.33 \cdot 10^{-5}$ )
	5 000	0.15	$1.21 \cdot 10^{-4}$ ( $9.98 \cdot 10^{-6}$ )	(0.15,0.05)	$1.21 \cdot 10^{-4}$ ( $9.88 \cdot 10^{-6}$ )
	10 000	0.15	$9.21 \cdot 10^{-5}$ ( $7.41 \cdot 10^{-6}$ )	(0.15,0.05)	$9.14 \cdot 10^{-5}$ ( $7.31 \cdot 10^{-6}$ )
			$\tilde{F}_n$		$\hat{F}_n^{MS}$
(0.4,0.4)	500	0.200	$5.56 \cdot 10^{-4}$ ( $4.56 \cdot 10^{-5}$ )	(0.250,0.250)	$7.21 \cdot 10^{-4}$ ( $5.58 \cdot 10^{-5}$ )
	1 000	0.100	$3.26 \cdot 10^{-4}$ ( $2.83 \cdot 10^{-5}$ )	(0.200,0.500)	$3.48 \cdot 10^{-4}$ ( $3.30 \cdot 10^{-5}$ )
	5 000	0.100	$1.10 \cdot 10^{-4}$ ( $9.98 \cdot 10^{-6}$ )	(0.200,0.333)	$7.20 \cdot 10^{-5}$ ( $7.11 \cdot 10^{-6}$ )
	10 000	0.067	$6.38 \cdot 10^{-5}$ ( $4.82 \cdot 10^{-6}$ )	(0.167,0.333)	$7.35 \cdot 10^{-5}$ ( $6.45 \cdot 10^{-6}$ )
(0.6,0.6)	500	0.200	$1.45 \cdot 10^{-3}$ ( $1.35 \cdot 10^{-4}$ )	(0.250,0.250)	$5.51 \cdot 10^{-4}$ ( $5.28 \cdot 10^{-5}$ )
	1 000	0.250	$3.59 \cdot 10^{-3}$ ( $1.97 \cdot 10^{-4}$ )	(0.250,0.200)	$4.13 \cdot 10^{-4}$ ( $3.40 \cdot 10^{-5}$ )
	5 000	0.333	$1.54 \cdot 10^{-2}$ ( $2.03 \cdot 10^{-4}$ )	(0.250,0.167)	$2.24 \cdot 10^{-4}$ ( $5.66 \cdot 10^{-5}$ )
	10 000	0.333	$1.50 \cdot 10^{-2}$ ( $1.48 \cdot 10^{-4}$ )	(0.250,0.200)	$1.32 \cdot 10^{-4}$ ( $7.23 \cdot 10^{-6}$ )

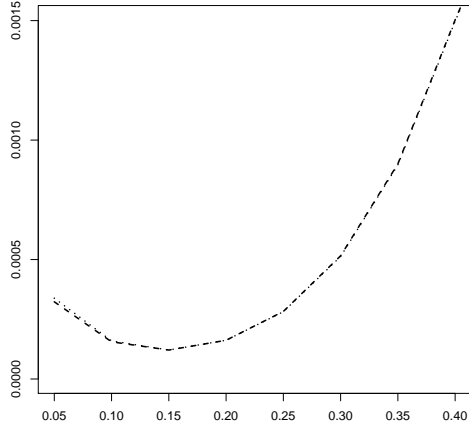
Table 1: Minimum values of the estimated MSE of the estimators  $\hat{F}_n^{(1)}$ ,  $\hat{F}_n^{(2)}$ ,  $\tilde{F}_n$  and  $\hat{F}_n^{MS}$  for different values of  $n$  at different points  $(t_0, z_0)$  for the simulation study. The values of  $\alpha_n$  and  $\beta_n$  that resulted in these minimal values are also given, as well as the standard errors of the mean of the squared differences between the estimator and the true value.



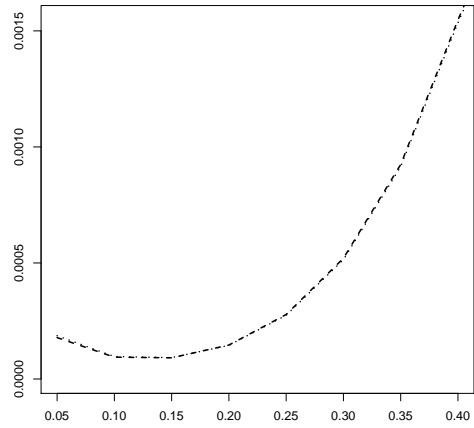
(a)  $n = 500$  with  $\beta_n = 0.15$  for  $\hat{F}_n^{(2)}$



(b)  $n = 1000$  with  $\beta_n = 0.05$  for  $\hat{F}_n^{(2)}$



(c)  $n = 5000$  with  $\beta_n = 0.05$  for  $\hat{F}_n^{(2)}$



(d)  $n = 10000$  with  $\beta_n = 0.05$  for  $\hat{F}_n^{(2)}$

Figure 5: The estimated MSE as function of the smoothing parameter  $\alpha_n$  the estimators  $\hat{F}_n^{(1)}$  (dotted line) and  $\hat{F}_n^{(2)}$  (dashed line) for different values of  $n$  at the point  $(0.6, 0.6)$  for the simulation study.

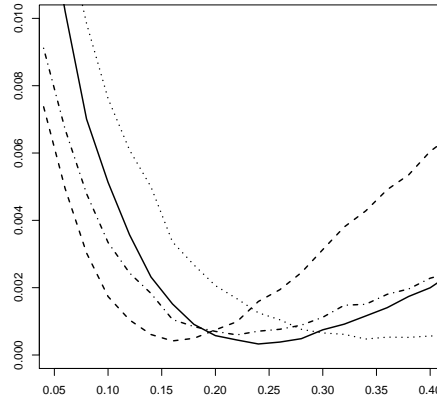


Figure 6: Values the estimated MSEs  $\widehat{MSE}^{(1)}(\alpha_n; 0.5, 0.5)$  (dotted line),  $\widehat{MSE}^{(2)}(\alpha_n, 0.78; 0.5, 0.5)$  (solid line),  $\widehat{MSE}^{(1)}(\alpha_n; 0.5, 0.5)$  (dash-dotted line) and  $\widehat{MSE}^{(2)}(\alpha_n, 0.8; 0.5, 0.5)$  (dashed line) as function of  $\alpha_n$ .